numero 226
giugno 1987
anno xx
volume XXXVIII

LE SCIENZE
edizione italiana di SCIENTIFIC AMERICAN

Armi nucleari della terza generazione

A differenza delle armi già schierate, le quali liberano l'energia esplosiva in modo indiscriminato, quelle future potrebbero essere in grado di produrre solo certi tipi di energia e concentrarli selettivamente sui bersagli

di Theodore B. Taylor

All'inizio degli anni cinquanta i laboratori militari degli Stati Uniti furono estremamente produttivi. Non solo ottennero miglioramenti spettacolari nel rendimento delle bombe a fissione (la prima generazione di armi nucleari), ma riuscirono anche a realizzare una seconda generazione di armi nucleari sfruttando la potenza esplosiva della fusione nella bomba a idrogeno e nei suoi vari derivati. Alla fine di quello stesso decennio le testate degli arsenali nucleari degli Stati Uniti assomigliavano ormai ben poco alle prime bombe che avevano aperto l'era nucleare, quelle sganciate su Hiroshima e Nagasaki.

Oggi è tecnologicamente possibile una terza generazione di armi nucleari. Modificando la forma dell'esplosivo nucleare e manipolando altre caratteristiche di progetto, si potrebbero costruire armi capaci di generare e dirigere fasci di radiazioni o getti di sferette metalliche su certi bersagli come basi missilistiche a terra, missili in volo e satelliti nello spazio. La travolgente ondata di creatività tecnica che produsse le prime due generazioni di armi nucleari può essere spiegata in larga misura dal fatto che i laboratori militari degli Stati Uniti potevano contare su fondi ingenti, su un mandato che li autorizzava a perseguire nuove possibilità in fatto di armi e sull'appoggio incondizionato del Governo. Inoltre, ponendosi nelle vesti di chi a quel tempo lavorava alla progettazione di armi nucleari, forse il fattore più stimolante era semplicemente costituito dall'euforia che prende ogni scienziato o ricercatore quando ha la libertà di esplorare concetti tecnici del tutto nuovi e poi di realizzarli.

La Strategic Defense Initiative (SDI), nell'ambito della quale è in corso un intenso programma militare di ricerca e sviluppo, potrebbe creare nei laboratori condizioni analoghe a quelle degli anni cinquanta. E' probabile che le difficoltà tecniche implicite nell'iniziativa promossa dal presidente Reagan per mettere a punto un sistema di difesa contro i missili balistici stimolino i progettisti di oggi a prendere in considerazione tipi radicalmente nuovi di armi nucleari, a prescindere completamente dai concomitanti progressi nei sistemi di lancio e di comando e controllo.

Sarebbe logico basarsi per la progettazione di questo nuovo tipo di armi su quelle della prima e della seconda generazione che trasformano la massa in una grande quantità di energia, la cui distribuzione avviene con un andamento a simmetria pressoché sferica. Una nuova generazione di armi nucleari potrebbe intensificare o sopprimere selettivamente certi tipi di energia di tutta quella prodotta da un'esplosione nucleare. Inoltre, gli effetti letali di un determinato portatore di energia (come radiazione elettromagnetica, particelle subatomiche e materiale espulso) potrebbero essere incrementati trasformando l'andamento a simmetria sferica in un andamento estremamente asimmetrico, tale da concentrare l'energia in una certa direzione.

In realtà, armi nucleari che riversino su un bersaglio una quantità di energia

«SFERA DI FUOCO»
NEUTRONI
RAGGI GAMMA
ONDA D'URTO
MICROONDE
LUCE VISIBILE E INFRAROSSA

L'andamento dell'emissione di energia distingue le testate nucleari attuali da quelle che potrebbero essere sviluppate tra breve. Le testate attuali (*in alto*) liberano l'energia esplosiva in molte forme, ognuna caratterizzata da irraggiamento uniforme verso l'esterno. Pertanto la zona in cui le installazioni militari verrebbero distrutte o messe fuori uso da ciascuno dei principali tipi di energia (*codice dei colori qui sopra*) può essere rappresentata all'incirca come una sfera. Le testate delle future armi nucleari potrebbero invece essere munite di dispositivi che eliminino, convertano e dirigano l'energia, trasformando una buona parte dell'energia esplosiva in microonde che sono poi concentrate sui bersagli (*in basso*).

per unità di superficie 1000 e più volte superiore a quella di un'arma nucleare convenzionale sono del tutto plausibili. Componenti o materiali speciali fissati all'esterno di un ordigno nucleare potrebbero convertire in una forma diversa l'energia liberata dalla detonazione; una particolare configurazione dell'esplosivo nucleare e del suo involucro potrebbe incanalare in certe direzioni la maggior parte dell'energia. Oppure l'energia liberata da un'esplosione nucleare potrebbe essere convertita e diretta, sfruttando l'effetto che un'esplosione di questo genere ha sull'ambiente. Se queste armi verranno costruite, indipendentemente dal loro scopo originario, verranno certo modificate per poter essere impiegate in un'ampia varietà di missioni tattiche e strategiche - offensive oltre che difensive - in ambienti di ogni genere.

Le armi della nuova generazione, al pari di quelle delle precedenti generazioni, potrebbero trarre la loro enorme energia esplosiva dalla fissione (la scissione di un nucleo da parte di un neutrone con la produzione di due nuovi nuclei) o da una combinazione di fissione e fusione (l'unione di due nuclei leggeri in un nucleo più pesante). Le esplosioni di fissione sono più facili da produrre e sostanzialmente equivalgono ad ammassare, nell'arco di circa un microsecondo, una sufficiente quantità di materiale fissile (come uranio 235 o plutonio 239) in un volume abbastanza piccolo da far sì che si generi rapidamente un enorme numero di neutroni adatti alla fissione. La rapida concentrazione di materiale fissile viene ottenuta facendo opportunamente detonare cariche di esplosivo convenzionale in modo da ammassare le varie frazioni di materiale in un'unica massa compatta.

Avviare un'esplosione di fusione è molto più complesso, in quanto sono necessarie temperature estremamente elevate (dell'ordine di centinaia di milioni di kelvin). In effetti, l'unico meccanismo pratico con cui produrre tali temperature in un ordigno trasportabile è un esplosivo a fissione. A quanto risulta, non è stato ancora progettato un esplosivo a fusione pura - senza ricorrere a un innesco a fissione.

Le reazioni di fusione non solo liberano, per unità di peso, molta più energia delle reazioni di fissione, ma producono anche un maggior numero di neutroni di alta energia. I neutroni addizionali possono in realtà «sovralimentare» la potenza esplosiva di un'arma a fissione se vengono fatti interagire con uranio o con plutonio nel «nucleo» dell'arma. Ponendo perciò in un'arma a fissione piccole quantità di combustibile termonucleare come trizio o deuterio (entrambi isotopi dell'idrogeno) si aumenta il rapporto complessivo tra la potenza esplosiva e il peso dell'arma, in quanto il peso aggiunto per la sovralimentazione è minimo.

A differenza delle armi sovralimenta-

La serie di effetti elencati nella legenda a sinistra potrebbe essere sfruttata militarmente dalla prossima generazione di armi nucleari che dovrebbe sopprimere alcuni effetti e intensificarne altri e forse anche canalizzarli in certe direzioni. Nello spazio (*riga in alto*) le armi nucleari potrebbero emettere raggi X incoerenti in tutte le direzioni (*a*) o raggi X coerenti in una direzione particolare (*b*). Le microonde possono attraversare facilmente l'atmosfera e potrebbero quindi raggiungere dallo spazio la superficie della Terra, specie se fossero concentrate (*c*). Anche i raggi gamma percorrono una certa distanza attraverso l'aria e potrebbero essere diretti verso obiettivi nell'alta atmosfera (*d*). I detriti ionizzati dell'arma prodotti da un'esplosione nucleare al di sopra dell'atmosfera, ma entro il campo magnetico della Terra, potrebbero produrre un potente impulso di radiazione elettromagnetica di grande lunghezza d'onda a mano a mano che distorcono le linee di forza del campo (*e*). Un effetto analogo si può ottenere nell'atmosfera (*riga al centro*): i raggi X potrebbero strappare elettroni alle molecole dell'aria, e creare in questa un improvviso sbalzo di corrente, il quale avrebbe come conseguenza l'emissione dell'impulso di onde radio (*f*). Potrebbero essere intensificati diversi effetti prodotti dalle armi nucleari come l'emissione di neutroni (*g*), l'onda d'urto (*h*) e l'ignizione di materiali combustibili (*i*). Obiettivi nello spazio potrebbero essere attaccati da microonde irradiate verso l'alto (*j*). L'energia delle esplosioni nel sottosuolo (*riga in basso*) potrebbe interagire fortemente con il mezzo circostante, tanto da produrre onde d'urto intensificate al suolo (*k*) o nell'acqua (*l*). Sarebbe possibile controllare la quantità e la distribuzione della ricaduta radioattiva delle armi nucleari, a seconda sia dei materiali scelti per rivestire l'arma stessa, sia del fatto che questa venga fatta esplodere nel sottosuolo (*m*) o in acqua (*n*). Infine, esiste la possibilità che l'onda di pressione di un'esplosione sotterranea espella proiettili ultraveloci attraverso una sorta di cannone nello spazio (*o*).

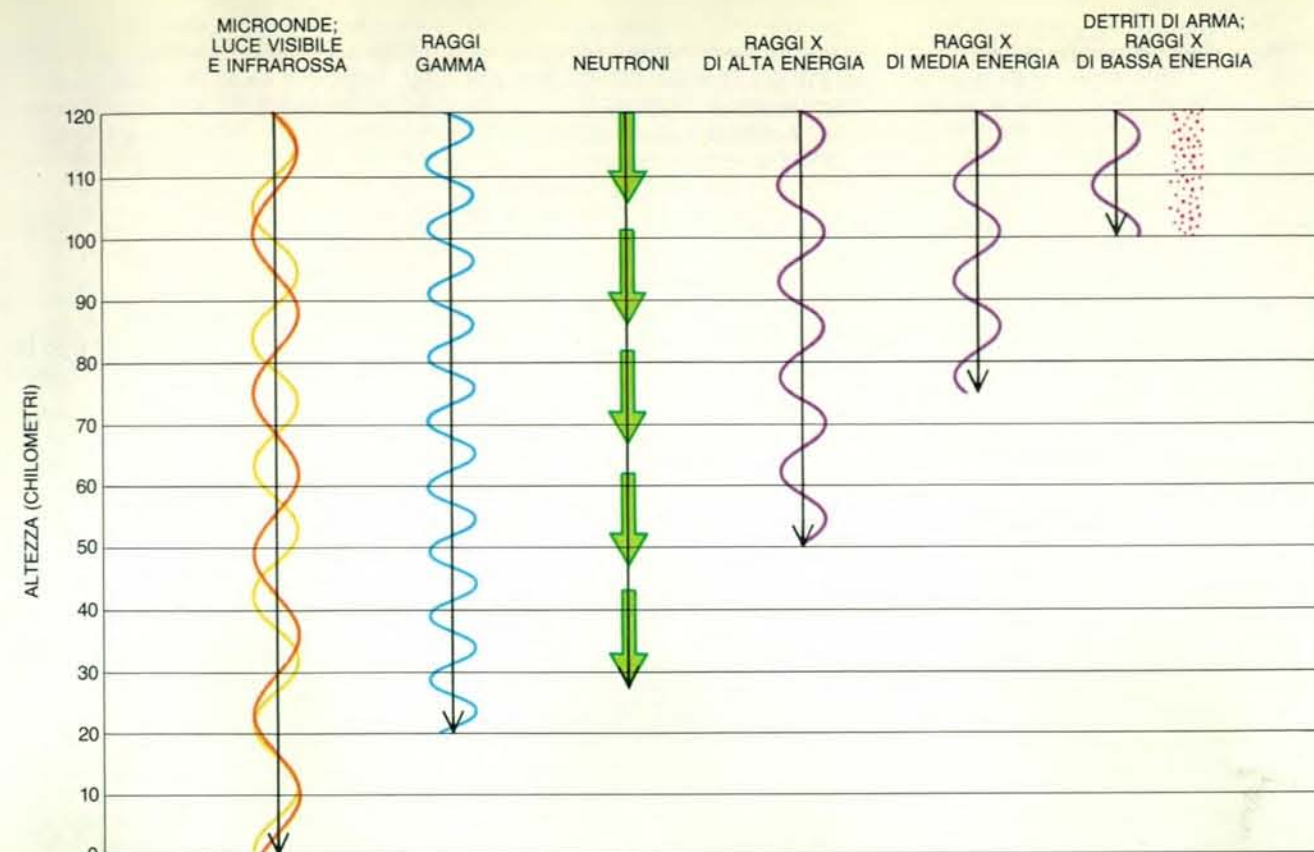
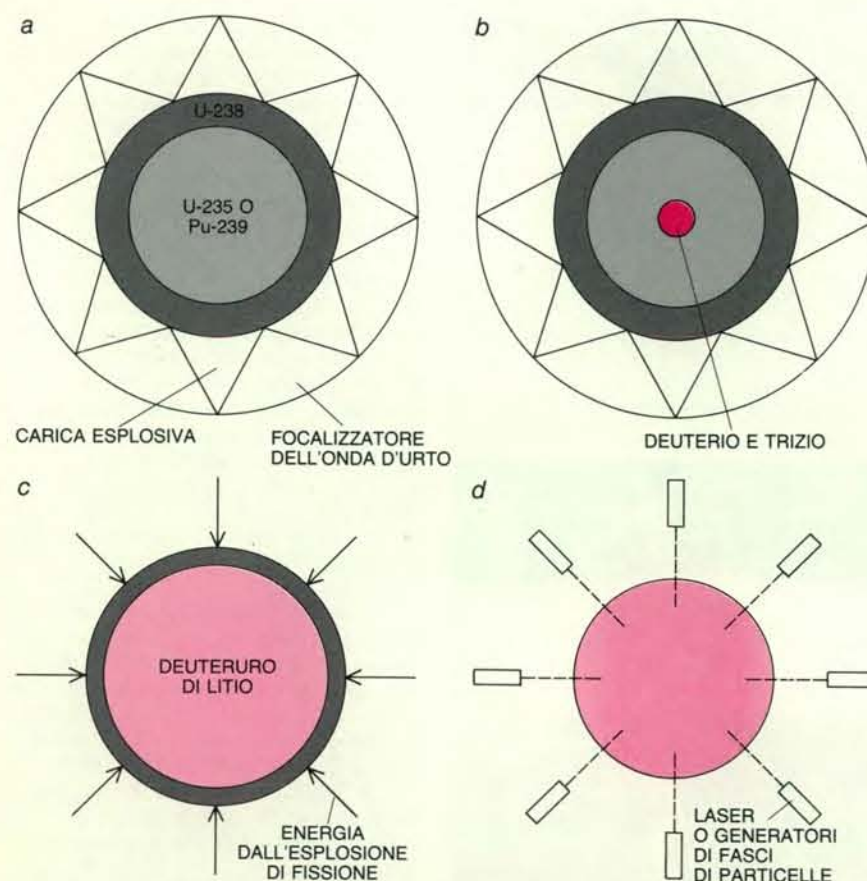
te, in cui l'energia liberata dalla fusione non dà un contributo significativo alla potenza esplosiva globale dell'arma, le cosiddette armi termonucleari traggono una parte notevole della loro energia esplosiva da reazioni di fusione. Le quantità relative di energia attribuibili alla fusione e alla fissione dipendono dal progetto dell'arma. Se una considerevole quantità di deuteruro di litio (che, quando viene bombardato con neutroni, produce trizio) viene compressa e riscaldata dall'energia liberata da un piccolo innesco a fissione, la frazione della potenza esplosiva totale dovuta alla fusione rispetto a quella dovuta alla fissione può diventare molto grande. Le armi di questo genere vengono chiamate a volte armi termonucleari «pulite», in quanto liberano una quantità relativamente ridotta di prodotti di fissione radioattivi.

All'altro estremo vi sono le armi in cui il combustibile termonucleare è racchiuso in una quantità considerevole di uranio naturale (uranio 238). I neutroni di alta energia prodotti dalla fusione in un combustibile termonucleare possono indurre la fissione nell'uranio circostante, moltiplicando notevolmente la potenza esplosiva totale della fissione.

Nelle testate a fissione pura, il rapporto potenza esplosiva-peso oscilla da un minimo di circa 0,0005 chiloton a un massimo di circa 0,1 chiloton per chilogrammo. Il rapporto globale potenza esplosiva-peso delle testate strategiche termonucleari è stato finora anche di circa sei chiloton per chilogrammo. Sebbene il massimo rapporto teorico per le reazioni di fissione e di fusione sia rispettivamente di 17 e di 50 chiloton per chilogrammo, per le armi degli Stati Uniti il massimo rapporto potenza esplosiva-peso si è avvicinato probabilmente al limite pratico per via di vari limiti inevitabili nella progettazione di armi nucleari (limiti dovuti innanzi tutto al fatto che è impossibile impedire all'arma di disintegrarsi prima che abbia avuto luogo la completa fissione dell'esplosivo nucleare). Ciononostante, il più basso rapporto tra potenza esplosiva e peso di un'arma a fissione pura è maggiore di parecchi ordini di grandezza di quello degli esplosivi chimici.

In realtà, la scarica di energia generata dall'esplosione di un'arma nucleare è tanto massiccia e violenta da vaporizzare e ionizzare l'arma stessa, convertendola in plasma, un gas estremamente caldo di ioni positivi e di elettroni. Inoltre, notevoli quantità di raggi gamma e di neutroni vengono emesse come prodotti secondari delle reazioni di fissione e di fusione. Tanto l'energia cinetica del plasma di detriti, quanto le emissioni nucleari costituiscono quelli che si potrebbero chiamare gli effetti primari di un'esplosione nucleare; essi si presentano in ogni esplosione di questa natura, indipendentemente dall'ambiente in cui essa ha luogo.

Alle temperature predominanti immediatamente dopo un'esplosione nucleare, il plasma irradia raggi X. In realtà, circa il 70 per cento dell'energia emessa nei primi microsecondi successivi a un'esplosione è costituito da raggi X. La frazione esatta dell'energia esplosiva totale liberata sotto forma di raggi X primari tende ad aumentare con il rapporto potenza esplosiva-peso, in quanto tale rapporto determina la temperatura globale del plasma di detriti. Quanto più grande è la quantità di energia dissipata in forma di raggi X, tanto minore è l'energia cinetica del plasma di detriti in espansione. Per un'arma termonucleare con un alto rapporto potenza esplosiva-peso, la velocità tipica del plasma sarebbe di circa 1000 chilometri al secondo, un valore che rappresenta circa il 10 per cento dell'energia esplosiva totale.

I raggi gamma che vengono emessi nel giro di un secondo circa dall'esplosione (i cosiddetti raggi gamma istantanei, o pronti) costituiscono circa il 3,5 per cento dell'energia totale liberata dalla fissione e ben il 20 per cento dell'energia liberata da alcuni cicli di reazioni termonucleari. Negli attuali tipi di esplosivi nucleari solo una piccola percentuale di questi raggi gamma viene assorbita all'interno dell'arma. L'energia cinetica dei neutroni eccedenti costituisce circa un altro 1,8 per cento dell'energia liberata dalla fissione e, a seconda del tipo di combustibile termonucleare, tra il 40 e l'80 per cento dell'energia liberata dalla fusione. I neutroni di alta energia tendono per altro a essere rallentati dalla diffusione anelastica dovuta a urti con elementi leggeri dei materiali dei sistemi di implosione. L'energia media dei neutroni che sfuggono realmente alla cattura nei materiali dell'arma e vengono liberati nell'ambiente circostante è quindi in genere molto inferiore. Questo effetto è particolarmente accentuato nelle armi termonucleari, in quanto il combustibile è costituito da elementi leggeri. In realtà, in armi di questo genere, l'energia dei neutroni viene deliberatamente depositata all'interno del combustibile termonucleare, poiché i neutroni hanno una funzione essenziale nel mantenere le elevate temperature necessarie ad alti tassi di reazione.

Gli esplosivi nucleari raffigurati si basano sulla fissione (la scissione di un nucleo pesante in due nuclei più leggeri per opera di un neutrone), a eccezione di uno. Un'arma la cui energia esplosiva è prodotta unicamente per fissione (*a*) è costituita da una parte centrale di materiale fissile (uranio 235 o plutonio 239) circondato da cariche di esplosivo convenzionale e da strutture inerti che concentrano verso l'interno l'energia dell'onda d'urto delle cariche, facendo implodere la parte centrale e dando così l'avvio a una reazione di fissione. Le armi basate sulla fissione possono essere «sovralimentate» (*b*) includendovi deuterio e trizio. Le temperature che si producono al momento della detonazione di un esplosivo a fissione fanno sì che gli isotopi dell'idrogeno subiscano la fusione dei nuclei, liberando grandi quantità di neutroni, i quali inducono altre reazioni di fissione. Nelle armi sovralimentate la reazione di fusione non contribuisce in misura significativa alla potenza esplosiva totale dell'arma. Le reazioni di fusione possono però rappresentare la maggior parte della potenza esplosiva di un'arma nucleare se una notevole quantità di un combustibile termonucleare come il deuteruro di litio viene esposta all'energia liberata dalla fissione (*c*). Un rivestimento esterno di uranio naturale (uranio 238) mantiene unita la testata per una frazione di microsecondo prima che essa si disintegri, permettendo alle reazioni nucleari di produrre più energia. Inoltre, quando viene irradiato dai neutroni prodotti dalla fusione, l'uranio subisce la fissione. Un'arma a fusione pura (*d*), che utilizzi fasci laser, ionici o elettronici per far implodere il combustibile termonucleare, sembra non sia stata ancora messa a punto.

La penetrazione in atmosfera dell'energia emessa da un'esplosione nucleare nello spazio dipende dal tipo di energia. Le radiazioni nella regione delle microonde, visibile e infrarossa dello spettro elettromagnetico potrebbero arrivare a terra con attenuazione piuttosto ridotta.

In questi ultimi 40 anni la maggior parte dello sviluppo delle armi nucleari non ha avuto altro scopo se non quello di intensificare o sopprimere in misura significativa particolari forme di energia, aggiustando le quantità relative di fissione e di fusione che hanno luogo nella testata. Un'eccezione è la cosiddetta bomba a neutroni (si veda l'articolo *La bomba N* di Fred M. Kaplan in «Le Scienze» n. 119, luglio 1978). Una bomba a neutroni è un esplosivo termonucleare di bassa potenza progettato per una maggior produzione di neutroni di alta energia per chiloton di potenza esplosiva totale. Nelle intenzioni, essa è un'arma nucleare destinata a neutralizzare l'esercito avversario.

Proprio come un'arma nucleare può essere progettata in modo da intensificare la sua emissione di neutroni primari a spese dell'onda d'urto e della ricaduta radioattiva, in teoria si potrebbe analogamente incrementare qualsiasi altro tipo di energia liberato da un esplosivo nucleare mettendo vicino all'esplosivo materiali appropriati disposti in opportune geometrie. Si potrebbe, per esempio, raggiungere un notevole controllo sulla quantità e sull'energia della radiazione X modificando il peso molecolare medio dei materiali dell'arma, la sua area superficiale esterna e il modo in cui l'energia generata nel suo «nucleo» viene distribuita sul fronte in espansione dei detriti dell'arma dopo la detonazione.

Anche modifiche di progetto potrebbero fare aumentare in misura significativa l'energia relativa dei raggi gamma istantanei. È possibile, per esempio, rivestire la carica nucleare con un «mantello» isotopico che, bombardato con neutroni, emetta raggi gamma. In questo modo i neutroni eccedenti della fissione o della fusione che sfuggono dal nucleo dell'arma potrebbero indurre l'emissione di raggi gamma, metà dei quali si allontanerebbe dai detriti in espansione dell'esplosione. (L'altra metà irradierebbe verso l'interno e verrebbe assorbita dal materiale detritico.)

Analogamente sarebbe possibile controllare, su un ampio raggio, la quantità dei prodotti radioattivi di fissione (la componente principale della ricaduta radioattiva) fra i detriti dell'arma, in particolar modo nel caso di armi termonucleari di potenza esplosiva superiore ad alcune centinaia di chiloton. Inoltre rivestendo l'arma con isotopi che, irradiati con neutroni, producano nuclei radioattivi con particolari periodi di dimezzamento e determinati modi di decadimento radioattivo, si potrebbe aumentare l'effetto letale della ricaduta radioattiva.

Gli effetti di un'esplosione nucleare potrebbero essere convogliati nella direzione voluta allo stesso modo in cui ordigni altamente esplosivi a carica profilata di tipo convenzionale possono produrre getti perforanti di metallo fuso o sciami di frammenti direzionali. Si possono trarre indicazioni utili dal modo in cui le cariche esplosive di forma non sferica liberano l'energia (*si veda l'illustrazione a pagina 25*).

Per esempio, facendo detonare tutto in una volta un disco di alto esplosivo, i prodotti dell'esplosione vengono scagliati fuori in una configurazione caratteristica a doppio cono. La ragione va ricercata nel fatto che la velocità dei prodotti dell'esplosione nella direzione perpendicolare alle due superfici del disco è maggiore della loro velocità radiale. Gli angoli al vertice dei coni saranno determinati dal rapporto fra lo spessore del disco e il suo diametro. Il flusso totale medio di energia cinetica (energia per unità di superficie per unità di tempo) dei prodotti dell'esplosione che attraversano un piano perpendicolare all'asse del doppio cono potrebbe avere un valore notevolmente maggiore di quello che avrebbe se la stessa massa di alto esplosivo espellesse i prodotti con un andamento a simmetria sferica. Se la velocità media dei prodotti dell'esplosione nella direzione dell'asse del cono equivalesse a 40 volte la loro velocità media nella direzione radiale (corrispondente a un angolo al vertice del cono di circa tre gradi), il fattore di intensificazione sarebbe circa 3000.

Nella detonazione di un cilindro lungo e sottile di alto esplosivo, si osserva invece che le velocità massime dei frammenti sono perpendicolari all'asse del cilindro. Perciò i prodotti dell'esplosione tendono a conservare una configurazione cilindrica; il fattore di intensificazione del flusso di energia è minore che nell'esempio precedente.

Infine, nel caso di una carica di alto esplosivo confinata da materiale denso in tutte le direzioni tranne che in avanti, i prodotti dell'esplosione vengono proiettati soprattutto in avanti. Il peso supplementare dovuto alla massa inerte intorno all'esplosivo è più che compensato dalla concentrazione dell'energia attraverso l'apertura.

Naturalmente, in un'esplosione nucleare viene liberata energia in molte forme diverse, quali raggi gamma, raggi X, neutroni e varie specie di radionuclidi, con intensità molto maggiori di quelle in gioco nella detonazione di un alto esplosivo convenzionale. E' ovvio che esplosivi nucleari anche di bassissima potenza offrono molte più opportunità degli esplosivi convenzionali di convogliare energia in una direzione voluta.

Molti degli effetti letali di un'esplosione nucleare sono in realtà risultati secondari dell'interazione dell'energia cinetica del plasma di detriti e della radiazione iniziale (raggi X) con il mezzo in cui ha luogo l'esplosione. Per questa ragione, molti fenomeni di interesse militare di un'esplosione nucleare sono determinati da proprietà del mezzo come pressione, densità e composizione. Le variazioni che si riscontrano in queste proprietà spiegano perché esplosioni nucleari nello spazio, nell'atmosfera, a livello del suolo e nel sottosuolo producano risultati assai diversi. Determinando, a seconda delle condizioni ambientali prevalenti, quali siano gli appropriati effetti primari da intensificare o da eliminare, è possibile trasferire più efficacemente ai bersagli gli effetti secondari dell'esplosione.

Dato che nello spazio non vi possono essere interazioni con alcun mezzo, i prodotti primari di un'esplosione nucleare (raggi X, plasma di detriti e radiazione nucleare) continuano a propagarsi nella stessa direzione in cui sono stati liberati finché non colpiscono qualcosa o non vengono deflessi dal campo magnetico o gravitazionale della Terra (a seconda che abbiano, rispettivamente, carica elettrica oppure massa). Di conseguenza, le asimmetrie iniziali nella distribuzione di massa di un esplosivo fatto detonare nello spazio tendono a conservarsi per grandi distanze nella configurazione dell'energia irradiata.

Se un esplosione nucleare avviene al di sopra dell'atmosfera, ma entro il campo magnetico della Terra, il plasma che si espande in direzioni più o meno perpendicolari alle linee di forza distorce tale campo. Quando questo avviene, buona parte dell'energia cinetica dei detriti viene convertita in energia elettromagnetica, e il risultato è l'emissione di un improvviso impulso di radiazione con un'ampia gamma di lunghezze d'onda, da pochi metri a centinaia di chilometri o più. L'impulso elettromagnetico (EMP) può costituire una parte notevole dell'energia totale dell'esplosione e può propagarsi nell'atmosfera, senza subire forti attenuazioni, fino alla superficie terrestre.

Le esplosioni nucleari nello spazio o nell'alta atmosfera possono produrre un altro tipo di EMP. In questo caso i raggi gamma o i raggi X di alta energia che colpiscono le regioni alte dell'atmosfera provocano l'espulsione di elettroni dalle molecole dell'aria. L'improvvisa cascata di elettroni è equivalente a un fortissimo impulso di corrente elettrica con una configurazione a simmetria non sferica (essa fluirebbe prevalentemente nella direzione in cui la densità dell'aria è maggiore, vale a dire verso il basso) e variabile nel tempo. Tale corrente genererebbe campi magnetici transitori che a loro volta produrrebbero radiazioni elettromagnetiche sotto forma di EMP.

Poiché la densità dell'atmosfera aumenta quasi esponenzialmente scendendo di quota, buona parte dell'energia irradiata verso il basso da un'esplosione nucleare che avviene nello spazio si accumula nell'alta atmosfera. L'accumulo di energia può produrre gravi effetti secondari che si propagano poi fino alla superficie terrestre. A fluenze (energia totale per unità di superficie) sufficientemente elevate, i raggi X e i detriti possono, per esempio, riscaldare l'atmosfera, portandola a temperature tanto elevate da far sì che essa emetta radiazioni infrarosse e visibili. Anche i raggi gamma, i neutroni e i raggi X liberati dall'esplosione, come i prodotti di decadimento dei radionuclidi, possono generare, direttamente o indirettamente, correnti elettriche nello strato dell'atmosfera nel quale depositano la loro energia. Queste correnti sono all'origine di altri EMP con una gamma molto ampia di lunghezze d'onda e di livelli istananei di potenza. Il riscaldamento dell'atmosfera può anche dare l'avvio a complesse reazioni chimiche che influenzano la trasmissione e la riflessione delle onde radio.

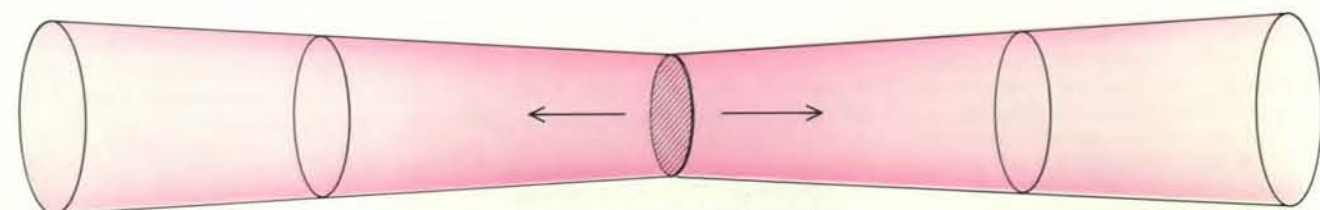
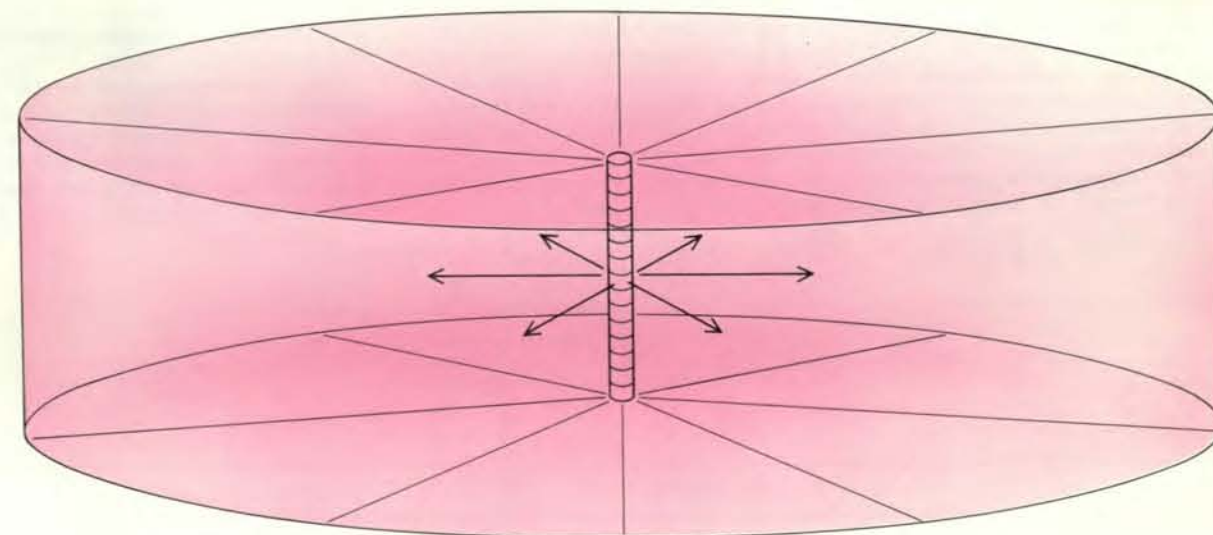
Nella bassa atmosfera, nel sottosuolo o in acqua la radiazione X primaria emessa in un'esplosione nucleare viene assorbita dagli atomi e dalle molecole del mezzo circostante entro qualche metro dal punto della detonazione. Di conseguenza, il mezzo si riscalda rapidamente, formando una «sfera di fuoco», che a sua volta riemette radiazioni elettromagnetiche di frequenza più bassa. La maggior parte di queste radiazioni è compresa nelle regioni visibile e infrarossa dello spettro e può percorrere nell'aria notevoli distanze.

L'energia radiante si combina anche con l'energia cinetica del plasma in espansione producendo sul mezzo circostante un impulso di pressione di forza tremenda. L'impulso genera un'onda d'urto, o di pressione, che si propaga nel mezzo; quanto più questo è denso, tanto maggiore è la quantità di energia che si trasforma in onda d'urto. Per questa ragione, nel caso di esplosioni in acqua o al suolo viene convertita in onda d'urto una percentuale più elevata di energia di quanto si verifichi nel caso di esplosioni in aria.

Le esplosioni al suolo, nel sottosuolo o a bassissima quota possono scagliare nell'aria anche enormi quantità di polvere, frammenti di terreno, acqua o parti di edifici e strutture che, direttamente o indirettamente, possono causare estese distruzioni. Per di più, è probabile che buona parte di questo materiale diventi radioattivo, contaminando gravemente vaste superfici per ricaduta.

In un'esplosione nucleare possono essere generate anche forme di energia che normalmente non vengono emesse come effetti primari o secondari e che si prestano a venire convogliate in fasci relativamente ristretti. Il problema fondamentale in questo caso (non possiamo parlarne dettagliatamente in questa sede perché l'argomento è coperto da segreto) è come convertire buona parte dell'energia di un'esplosione nucleare in una forma di energia che possa essere emessa in direzioni ben determinate. Basti dire che le radiazioni elettromagnetiche di lunghezze d'onda tipiche dei raggi gamma, dei raggi X, della luce visibile e delle microonde possono essere focalizzate con dispositivi analoghi al laser, che inducano atomi e molecole a irradiare in fase. Radiazioni di lunghezza d'onda superiore possono essere emesse nella direzione voluta se le armi in questione vengono dotate dell'equivalente di un'antenna. In entrambi i casi, il problema è come incanalare l'enorme flusso di energia di un'esplosione nucleare in un dispositivo che converta e convogli l'energia nei pochi microsecondi prima che l'intero complesso dell'arma si disintegri. Un'altra possibilità, che potrebbe forse semplificare un poco il problema, è di predisporre gli ordigni nucleari in una struttura di contenimento riutilizzabile dalla quale sia possibile poi attingere l'energia dell'esplosione. Tali strutture, progettate in modo da resistere a esplosioni di potenza pari anche a un chiloton, sono in realtà in fase di studio da parecchi decenni. Il Lawrence Livermore National Laboratory ha preso di recente in considerazione la proposta di costruire una camera di questo genere, in cui potrebbe essere studiato sperimentalmente un gran numero di effetti nucleari.

Per le armi con base a terra destinate ad attaccare obiettivi nello spazio il peso delle apparecchiature necessarie non è importante, mentre invece è un fattore critico per le armi con base nello spazio. E' quindi lecito attendersi che le tecnologie usate per sviluppare armi nucleari a energia orientata con base a terra saranno diverse da quelle richieste per armi analoghe con base nello spazio. Da un punto di vista militare, le armi con base a terra presentano numerosi vantaggi: la certezza di eludere eventuali trattati che mettano al bando le armi nucleari nello spazio, il facile accesso a grandi e pesanti apparecchiature di conversione (che permettano una più elevata capacità direzionale e una maggiore efficienza di conversione dell'energia dell'esplosione nella forma da emettere), il costo molto più basso e la possibilità di riutilizzo dell'apparecchiatura.

È possibile convertire l'energia dell'esplosione in impulsi elettrici per mezzo di generatori magnetoidrodinamici: dispositivi che convertono direttamente in elettricità l'energia cinetica di un plasma. (Tali dispositivi sono stati proposti per convertire in elettricità l'energia di fusione di un reattore di potenza.) Gli impulsi elettrici potrebbero allora azionare dei congegni per la conversione dell'elettricità in radiazione elettromagnetica (con o senza la conseguente autodistruzione del dispositivo) che potrebbe poi essere concentrata con grande precisione su un bersaglio nello spazio. Nella

ESPLOSIVO A FORMA DI DISCO PIATTO

ESPLOSIVO A FORMA DI CILINDRO LUNGO E SOTTILE

ESPLOSIVO CONFINATO

Cariche convenzionali di forma determinata possono emettere i prodotti dell'esplosione con simmetria non sferica. La configurazione caratteristica dei prodotti dell'esplosione di un disco piatto di esplosivo, per esempio, è un doppio cono. La detonazione di un esplosivo di forma cilindrica, lunga e sottile, produce un andamento a simmetria cilindrica. Infine, confinando con un rivestimento di materiale denso e inerte gli effetti dell'esplosione in tutte le direzioni tranne che in una, è possibile concentrare i prodotti dell'esplosione in quella direzione. Applicando effetti direzionali anche a esplosivi nucleari si potrà forse controllare l'andamento con cui vengono emessi i prodotti dell'esplosione.

maggior parte dei casi il basso rendimento di questa conversione di energia può essere più che compensato da un alto grado di focalizzazione in direzione del bersaglio.

Una possibilità più estrema è quella di sfruttare un'esplosione nucleare relativamente poco potente a grande profondità nel sottosuolo per accelerare proiettili attraverso l'equivalente di una canna di cannone. I cosiddetti proiettili ultraveloci raggiungerebbero velocità vicine a 10 chilometri al secondo, che è la velocità di fuga dalla Terra. Proiettili compatti di forma appropriata potrebbero perciò penetrare nell'atmosfera in maniera in qualche modo analoga a quanto fanno le grandi meteoriti. Una proposta simile fu studiata già negli anni cinquanta come metodo per portare nello spazio carichi pesanti a un costo relativamente basso.

Per esempio, l'energia cinetica di 10 tonnellate di materiale che si muovono alla velocità di 10 chilometri al secondo è equivalente all'energia dell'esplosione di circa 100 tonnellate di tritolo. Si può dedurre quindi che un'esplosione nucleare di potenza pari più o meno a un chiloton possa fornire energia propulsiva in misura più che sufficiente. Se la «canna del cannone» fosse lunga alcune centinaia di metri, l'accelerazione media del proiettile sarebbe all'incirca 10 000 volte l'accelerazione di gravità della Terra, che non è tale da superare la capacità di tolleranza alla deformazione di un proiettile compatto di alta densità. La successiva frammentazione in schegge solide o goccioline fuse potrebbe fare di un proiettile di questo genere un'arma di grande efficacia per distruggere satelliti o testate di missili balistici nello spazio.

Un'ulteriore possibilità è quella che la detonazione stessa acceleri il materiale che riveste l'arma, frammentandolo immediatamente in proiettili o goccioline che si muovono a velocità molto superiori a 10 chilometri al secondo. I frammenti ultraveloci così prodotti potrebbero essere concentrati in un volume conico, ma sarebbe necessario un meccanismo che controllasse il processo di accelerazione al fine di evitare la vaporizzazione dei frammenti. Inoltre armi di questo genere dovrebbero portare attacchi solamente contro obiettivi nello spazio o nell'alta atmosfera, in quanto a basse quote il raggio d'azione dei frammenti è in genere molto inferiore alla distanza alla quale l'onda d'urto provoca effetti distruttivi.

I danni che con tutta probabilità un bersaglio è destinato a subire quando è esposto a tutta la gamma dei tipi di energia liberati da un'esplosione nucleare possono essere calcolati approssimativamente valutando il tipo di energia che probabilmente raggiungerà l'oggetto, le modalità del danneggiamento e in molti casi anche la velocità di deposizione dell'energia. Questo aspetto degli effetti di un'esplosione nucleare è estremamente complesso e non ancora ben compreso.

La fluenza energetica totale in grado di provocare danni temporanei o permanenti a obiettivi militari o civili varia di nove ordini di grandezza (*si veda l'illustrazione in basso*). Gli effetti delle radiazioni di lunghezza d'onda maggiore (come quelle prodotte da un EMP), all'estremità inferiore della scala della fluenza energetica, sono i più difficili da valutare e rimangono quindi incerti.

La fluenza di 0,1 joule per metro quadrato è un milione di volte più grande di un normale segnale radio della durata di un secondo, emesso con simmetria sferica da un radiotrasmettitore da 10 chilowatt situato a 100 chilometri di distanza. È peraltro noto che le comunicazioni militari e commerciali e le trasmissioni radar, pur producendo fluenze assai più ridotte, hanno provocato in passato l'attivazione accidentale di detonatori di alti esplosivi e disturbi in calcolatori e altre apparecchiature elettriche ed elettroniche. Effetti analoghi sarebbero prodotti dall'EMP di un'esplosione nucleare. In realtà, sono già state prese misure per schermare apparecchi di interesse militare dagli effetti delle radiazioni elettromagnetiche; per esempio, si possono proteggere gli apparecchi in questione con schermature conduttrici ed evitare l'impiego di componenti che possano essere sensibili anche a piccoli impulsi di corrente indotti dalla radiazione elettromagnetica che riesce a penetrare. Non sempre, però, queste misure hanno avuto pieno successo.

Alcuni componenti di sistemi elettronici, come i transistori, possono essere molto sensibili a piccole correnti e altri effetti dovuti a raggi gamma o a bombardamento di neutroni. Benché questi effetti possano essere ridotti al minimo con la schermatura, la generale mancanza di misure protettive rende i sistemi spaziali non militari particolarmente vulnerabili a tali radiazioni.

I raggi gamma, i neutroni, i raggi X di alta energia o i radionuclidi che colpiscono un obiettivo nello spazio possono fargli assumere un potenziale dell'ordine della massima energia delle particelle cariche emesse. E' possibile che l'intensità

TIPO DI ENERGIA
MICROONDE E RADIAZIONE DI MAGGIORE LUNGHEZZA D'ONDA — SEGNALI ELETTRONICI SPURI; SOVRACCARICHI DI CORRENTE; AZIONAMENTO AUTODISTRUTTIVO DEI DETONATORI
RAGGI GAMMA E NEUTRONI DI ALTA ENERGIA — SOVRACCARICHI DI CORRENTE; DOSE LETALE PER L'UOMO
FRAMMENTI DI ARMA ULTRAVELOCI — PENETRAZIONE DELLA SUPERFICIE ESTERNA; DISTACCO DELLA SUPERFICIE INTERNA
LUCE VISIBILE E INFRAROSSA — ACCENSIONE DI MATERIALE COMBUSTIBILE
PLASMA DI DETRITI — DISTACCO DELLA SUPERFICIE INTERNA; DANNI STRUTTURALI
RAGGI X — DISTACCO DELLA SUPERFICIE INTERNA; DANNI STRUTTURALI
10^{-1} 1 10 10^2 10^3 10^4 10^5 10^6 10^7 10^8 10^9
FLUENZA CHE PROVOCA DANNI (JOULE PER METRO QUADRATO)

La fluenza (energia totale per unità di superficie) necessaria per distruggere obiettivi militari può variare di nove ordini di grandezza, a seconda dei vari tipi di energia emessi in un'esplosione nucleare. È probabile che nella terza generazione di armi nucleari vengano sfruttate radiazioni con lunghezze d'onda corrispondenti alle microonde o superiori, che sono in grado di provocare danni anche a fluenze relativamente ridotte.

del campo elettrico vicino alla superficie raggiunga valori intorno a un milione di volt per metro, sufficienti per indurre danni temporanei o permanenti nei sistemi elettrici interni che non siano ben protetti.

A differenza dei neutroni e dei raggi gamma, i frammenti ultraveloci sono in grado di perforare la superficie di un bersaglio. L'espulsione estremamente rapida di materiale al momento della formazione dei fori produrrebbe nel bersaglio una forte onda d'urto. A causa della loro alta velocità, che supera anche di un centinaio di volte quella di un proiettile ultrarapido di fucile, frammenti di peso molto inferiore a un grammo possono provocare notevoli danni quando vengano diretti contro obiettivi nello spazio.

La radiazione infrarossa o visibile emessa come effetto secondario del riscaldamento atmosferico provoca danni soprattutto incendiando i materiali combustibili sulla superficie del bersaglio. Anche se la superficie non è combustibile, il riscaldamento non uniforme può provocare ugualmente danni in conseguenza delle sollecitazioni termiche che in tal modo si producono.

I raggi X di alta energia e il plasma di detriti possono danneggiare un bersaglio nello spazio soprattutto provocando la rapida vaporizzazione di materiale dalla superficie del bersaglio stesso. Se l'agente è costituito da raggi X, la forza dell'impatto può essere trasmessa attraverso gli strati esterni dell'oggetto, fino a frantumarne le superfici interne, ammesso che il tempo necessario per depositare l'energia incidente sia breve rispetto al tempo richiesto perché l'impatto raggiunga la superficie interna. Un processo del genere è detto *spalling* (distacco superficiale). Il plasma di detriti, invece, di solito non provoca fenomeni di distacco, essendo necessario un tempo troppo lungo perché l'energia cinetica del plasma venga ceduta. In ogni caso, la quantità di moto totale trasferita verso l'interno dall'evaporazione superficiale può mettere fuori uso il bersaglio, anche senza provocare fenomeni di distacco superficiale.

Per rendere più comprensibili queste stime, consideriamo la distanza alla quale effetti distruttivi possono essere prodotti da un determinato portatore di energia (*si veda l'illustrazione in questa pagina*). Effetti distruttivi su scala potenzialmente enorme (ovvero, una fluenza elevata a una data distanza) possono essere facilmente ottenuti emettendo energia entro un angolo ristretto. Le microonde di lunghezza d'onda compresa fra tre centimetri e un metro sono particolarmente adatte a un'intensificazione direzionale di questo genere perché in questa regione dello spettro l'atmosfera è sostanzialmente trasparente, permettendo l'impiego di questo tipo di radiazione in sistemi d'arma terra-spazio, spazio-terra e spazio-spazio. Inoltre, i valori di fluenza energetica necessari per provocare danni a molti tipi di obiettivi civili e militari sono più bassi nella regione delle microonde rispetto a tutti gli altri tipi di radiazione elettromagnetica.

TIPO DI ENERGIA
MICROONDE E RADIAZIONE DI MAGGIORE LUNGHEZZA D'ONDA — 10%
DANNI ALL'ELETTRONICA
RAGGI GAMMA E NEUTRONI DI ALTA ENERGIA — 20% — 20%
DOSE LETALE PER L'UOMO
FRAMMENTI DI ARMA ULTRAVELOCI — 10%
LUCE VISIBILE E INFRAROSSA — 10%
PLASMA DI DETRITI — 40%
RAGGI X — 80%
10^{-2} 10^{-1} 1 10 10^2 10^3 10^4 10^5
RAGGIO DEL DANNO (CHILOMETRI)

La distanza massima dal punto di un'esplosione nucleare entro cui obiettivi militari nello spazio vengono danneggiati dipende dal tipo di energia implicata e da quale frazione dell'energia esplosiva totale essa rappresenti. Vengono considerate esplosioni della potenza di un chiloton (*in nero*) e di un megaton (*in colore*). Le barre indicano il raggio stimato dei danni provocati da armi della terza generazione, a energia intensificata, ma non diretta. Accanto a ogni coppia di barre viene indicata la frazione di energia esplosiva totale convogliata in ogni particolare tipo di energia. Armi a energia diretta potrebbero provocare danni in un raggio molto più grande.

Il potenziale militare dei fasci di microonde diretti è spaventoso. Si supponga, per esempio, che sia possibile convertire il 5 per cento dell'energia liberata da un'esplosione di un chiloton in radiazione della lunghezza d'onda di tre centimetri emessa da un'antenna del diametro di 50 metri o dall'equivalente di un laser a microonde. L'esplosione di un ordigno di questo genere, in un'orbita geostazionaria all'altezza di 30 000 chilometri, depositerebbe circa 800 joule per metro quadrato su una superficie di 250 chilometri quadrati (più grande dell'area di Washington). La fluenza energetica prevista è superiore al livello che notoriamente provoca gravi danni a molti tipi di apparecchiature elettriche: calcolatori, antenne, relè e linee elettriche. Naturalmente a distanze molto più brevi la fluenza energetica sarebbe molto maggiore: circa cinque milioni di joule per metro quadrato a una distanza di 400 chilometri.

Lo sviluppo e lo spiegamento di un sistema d'arma a microonde complicherebbe notevolmente le tattiche e le strategie militari sia offensive, sia difensive. Potrebbe, per esempio, provocare danni temporanei o permanenti alle complesse apparecchiature elettriche ed elettroniche normalmente impiegate nei sistemi per la sorveglianza, l'inseguimento, le comunicazioni, la navigazione e altre funzioni di comando e controllo. Poiché l'atmosfera è virtualmente trasparente alle microonde, sia il dispositivo che genera il fascio, sia il bersaglio previsto potrebbero trovarsi nello spazio, nell'atmosfera o sulla superficie terrestre. In ogni caso, è probabile che lo spiegamento di tali armi possa minare alla base la fiducia nell'affidabilità delle forze tattiche e strategiche in caso di guerra, comprese quelle che costituiscono il deterrente principale alla guerra nucleare.

Quali sono le probabilità che questa terza generazione di armi nucleari venga effettivamente sviluppata e schierata? La risposta dipende in gran parte da come e in che misura l'Unione Sovietica e gli Stati Uniti sostengono l'attività di ricerca dei rispettivi laboratori militari. Poiché gli sviluppi in campo militare di un paese provocano inevitabilmente nell'altro paese risposte emulative, la probabilità dipende fortemente da come viene percepito l'andamento della ricerca e dello sviluppo dell'avversario in questo settore.

Un indice fondamentale delle dimensioni dello sforzo di un paese è la frequenza della sperimentazione nucleare. Se gli Stati Uniti continueranno e l'Unione Sovietica riprenderà gli esperimenti nucleari nel sottosuolo anche a livelli molto inferiori al limite di 150 chiloton fissato dal Threshold Test Ban Treaty, sarà probabilmente solo questione di tempo prima che vengano sviluppati questi nuovi tipi di armi nucleari offensive e difensive.

L'industria europea di minerali e metalli

Il futuro dell'industria mineraria e di quella metallurgica nella Comunità economica europea è condizionato in maniera determinante dalla scarsità delle risorse naturali di metalli e dalla carenza di fonti energetiche

di Felice Ippolito e Barbara Curli

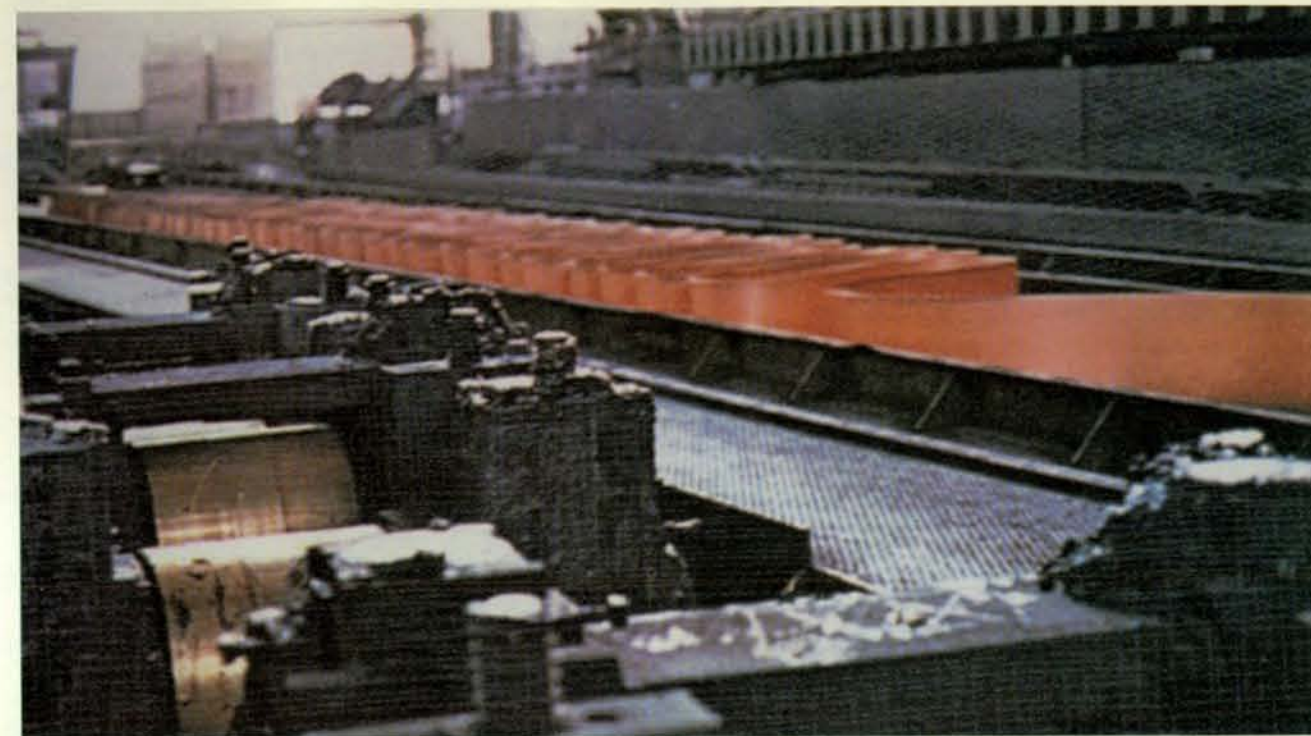
In Europa, e in particolar modo nella Comunità economica europea (CEE), il problema dell'approvvigionamento di materie prime per uso industriale è stato alquanto trascurato negli ultimi anni, in quanto le gravi crisi petrolifere degli anni settanta concentravano l'attenzione degli studiosi e del pubblico, e condizionavano le scelte politiche, prevalentemente sulle fonti energetiche. Eppure il settore minerario non è meno importante per il futuro industriale delle economie europee, poiché l'Europa è carente di risorse minerarie metallifere e dipende largamente dall'estero per gli approvvigionamenti primari di taluni elementi indispensabili allo sviluppo industriale.

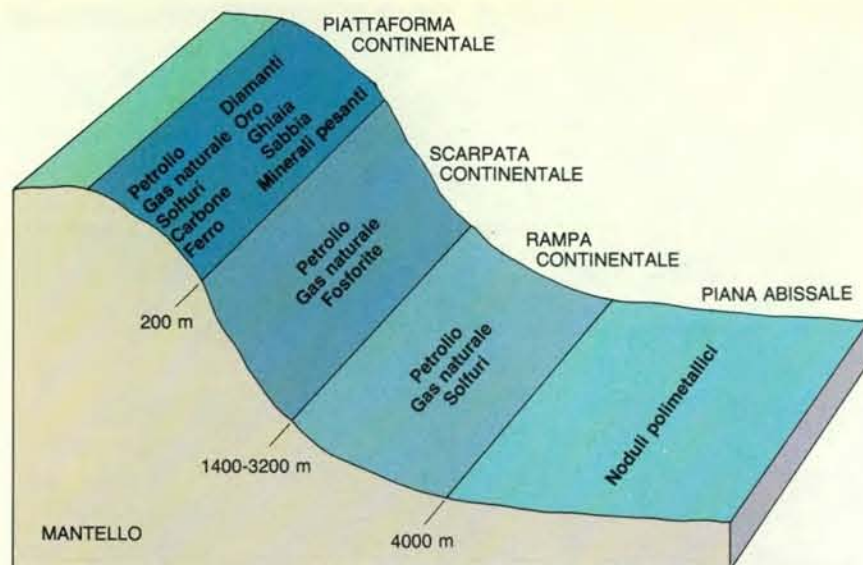
A questa questione di fondo è strettamente collegata la grave crisi produttiva e occupazionale in cui versa l'industria europea di prima trasformazione; una crisi da sovraproduzione dovuta alla stagnazione della domanda, in larga misura a sua volta conseguente alla maggiore concorrenza da parte dei paesi in via di industrializzazione, che sono ormai in grado di produrre ed esportare prodotti di prima lavorazione a prezzi inferiori rispetto a quelli europei. Inoltre, l'evoluzione dell'industria metallurgica non è scindibile dal problema energetico in quanto la Comunità europea non solo è povera di depositi minerali metalliferi, ma è anche dipendente dall'estero per oltre il 40 per cento dei suoi fabbisogni energetici. Essa pertanto è costretta a importare sia le materie prime, sia l'energia occorrente per la loro lavorazione primaria, con la conseguenza di gravosi costi addizionali.

Il futuro del settore minerario e metallurgico dipende strettamente anche dai mutamenti attualmente in corso nel campo delle nuove tecnologie e dei materiali avanzati, che stanno già modificando la composizione relativa della domanda di minerali e metalli, privilegiandone taluni rispetto ad altri. Questo avviene sulla base di un processo di evoluzione e di sostituzione di materiali caratteristico dell'attuale fase di sviluppo tecnologico avanzato.

Infine, occorre considerare il problema della tutela ambientale verso il quale oggi l'opinione pubblica è così sensibile: infatti, l'industria metallurgica è altamente inquinante, onde si impone non solo un riesame di tutti gli impianti esistenti ai fini di diminuire il loro impatto sull'ambiente, ma anche una analisi preliminare di costi/benefici per le nuove installazioni. L'attuale temporanea favorevole congiuntura internazionale nel settore energetico non dovrebbe infatti invitare a sottovalutare aspetti finora passati in secondo piano, o ai quali fino a questo momento si sono riservate soluzioni vecchie e di tipo congiunturale. Tali problemi sono comunque assai importanti per lo sviluppo economico e per la strategia energetica degli anni futuri, la cui programmazione richiede tempi lunghi e scelte ragionate di politica industriale, oltre che una visione globale europea delle linee di intervento.

La Commissione della CEE e il Parlamento europeo si sono a più riprese interessati al problema delle materie prime minerarie. Ricordiamo, per citare gli interventi più recenti, la relazione Moreau (gennaio 1982) sul bilancio e le prospettive di approvvigionamento dell'Europa comunitaria per quanto riguarda materie prime minerali e vegetali; la decisione del Consiglio dei ministri della Comunità (maggio 1982) che adottava un programma triennale di ricerca e sviluppo nel settore delle materie prime; la proposta della Commissione al Consiglio (luglio 1985) per un programma di ricerca sulle materie prime e i materiali avanzati per gli anni 1986-1989; infine, la relazione al Parlamento europeo, che uno di noi presentò (maggio 1985) sulla valorizzazione delle risorse del sottosuolo comunitario. Finora, tuttavia, ben poco è stato fatto in concreto per affrontare i nodi del problema, segnatamente per la crisi di bilancio in cui si dibatte la Comunità. Manca di conseguenza una impostazione strategica di fondo, economica non meno che politica, che da un canto colleghi le prospettive dell'approvvigionamento primario ai problemi energetici e alla situazione dell'industria di trasformazione europea e dall'altro tenga conto delle tendenze del mercato internazionale e dei prevedibili sviluppi tecnologici.

È ben nota da tempo la povertà di materie prime minerarie dei paesi membri della Comunità. Considerando i 15 principali minerali e metalli per uso industriale (*si veda l'illustrazione in basso a pagina 30*), la produzione primaria consente un grado di copertura dei consumi (autosufficienza primaria) che va dallo 0 al 10 per cento per 12 di essi e dal 15 al 30 per cento per soli tre di essi (alluminio, zinco e tungsteno). Questa situazione non è sostanzialmente mutata con l'adesione di Spagna e Portogallo, se non per un lieve maggiore apporto alla produzione totale di ferro, rame e zinco. Anche se la produzione secondaria interna (riciclaggio) permette un sensibile miglioramento del grado di autosufficienza, la dipendenza dalle importazioni extracomunitarie continua a oscillare tra il 40 e il 90 per cento e oltre dei consumi delle materie prime considerate. Tale dipendenza è quindi del tutto analoga, in termini quantitativi, a quella di fonti energetiche e segnatamente di idrocarburi. Ma a parte gli aspetti quantitativi, già di per sé eloquenti nel fornire il qua-

Mentre, dal 1974 al 1983, gli addetti alla siderurgia sono diminuiti nella Comunità europea di oltre il 40 per cento, in Italia l'occupazione si è ridotta meno che negli altri paesi a causa di massicci investimenti pubblici effettuati nel corso degli anni settanta in molti stabilimenti, come per esempio quello di Bagnoli, del gruppo Italsider, ripreso nella fotografia. Gli effetti negativi di questi interventi si stanno facendo sentire, oltre che nelle pesanti perdite economiche, anche nelle maggiori difficoltà che si incontrano nell'ammodernamento degli impianti del settore.

L'intero territorio dell'Antartide è dal 1959 sotto la giurisdizione di un trattato internazionale, il cosiddetto «Trattato antartico», che è stato firmato anche dall'Italia nel 1981. Per il 1991 è previsto un rinegoziato cui potranno partecipare tutti i paesi che per quella data avranno stabilito in quel continente una base fissa di studi e di esplorazioni. La stazione italiana (*nella fotografia*) è stata definitivamente installata nel 1986 nella Terra di Vittoria (*si veda l'illustrazione a pagina 31*). L'importanza della presenza italiana è giustificata dalle possibilità di rinvenire in Antartide risorse petrolifere e metallifere rilevanti ed economicamente sfruttabili. La Comunità europea come tale è per ora assente dall'iniziativa, mentre sarebbe auspicabile che i paesi membri avessero la possibilità di agire nell'ambito di un più vasto programma europeo.

Nel profilo delle terre sommerse è indicata la distribuzione delle maggiori risorse di minerali rinvenibili alle diverse profondità. I noduli polimetallici, che potrebbero costituire una fonte cospicua di metalli di impiego industriale, si trovano in mare aperto e a notevoli profondità. Il loro eventuale sfruttamento commerciale, che richiede tecnologie particolari, d'altra parte già in fase di avanzato sviluppo, non è stato ancora disciplinato giuridicamente a livello sovranazionale.

dro della situazione, è utile soffermarsi su taluni aspetti qualitativi dell'approvvigionamento primario.

A causa della crisi di sovraproduzione dell'industria siderurgica, la produzione primaria di ferro nei paesi della Comunità si è ridotta nel 1982 a quasi la metà del valore del 1978 (*si veda l'illustrazione in alto di pagina 32*). Nonostante tale contrazione, l'81 per cento delle importazioni totali di ferro della Comunità è ancora costituito da importazioni di minerale di ferro. Nei casi dello zinco e dell'alluminio, i cui mercati registrano analogamente un eccesso di capacità produttiva, il 75 per cento delle importazioni di zinco è costituito da minerale di zinco e oltre il 40 per cento di quelle di alluminio da bauxite e altri minerali di alluminio. Per quanto si possano già osservare lievi, e inevitabili, mutamenti in questa tendenza, sembra pertanto che la Comunità continui a privilegiare una «filosofia» tradizionale degli approvvigionamenti minerari, fondata sulle importazioni di prodotti primari da destinare alla trasformazione di base, e che pertanto non tenga nella dovuta considerazione importanti fattori di medio-lungo periodo quali la concorrenza internazionale, l'industrializzazione nei paesi in via di sviluppo, la situazione economica dell'industria di trasformazione e le tendenze generali di mercato.

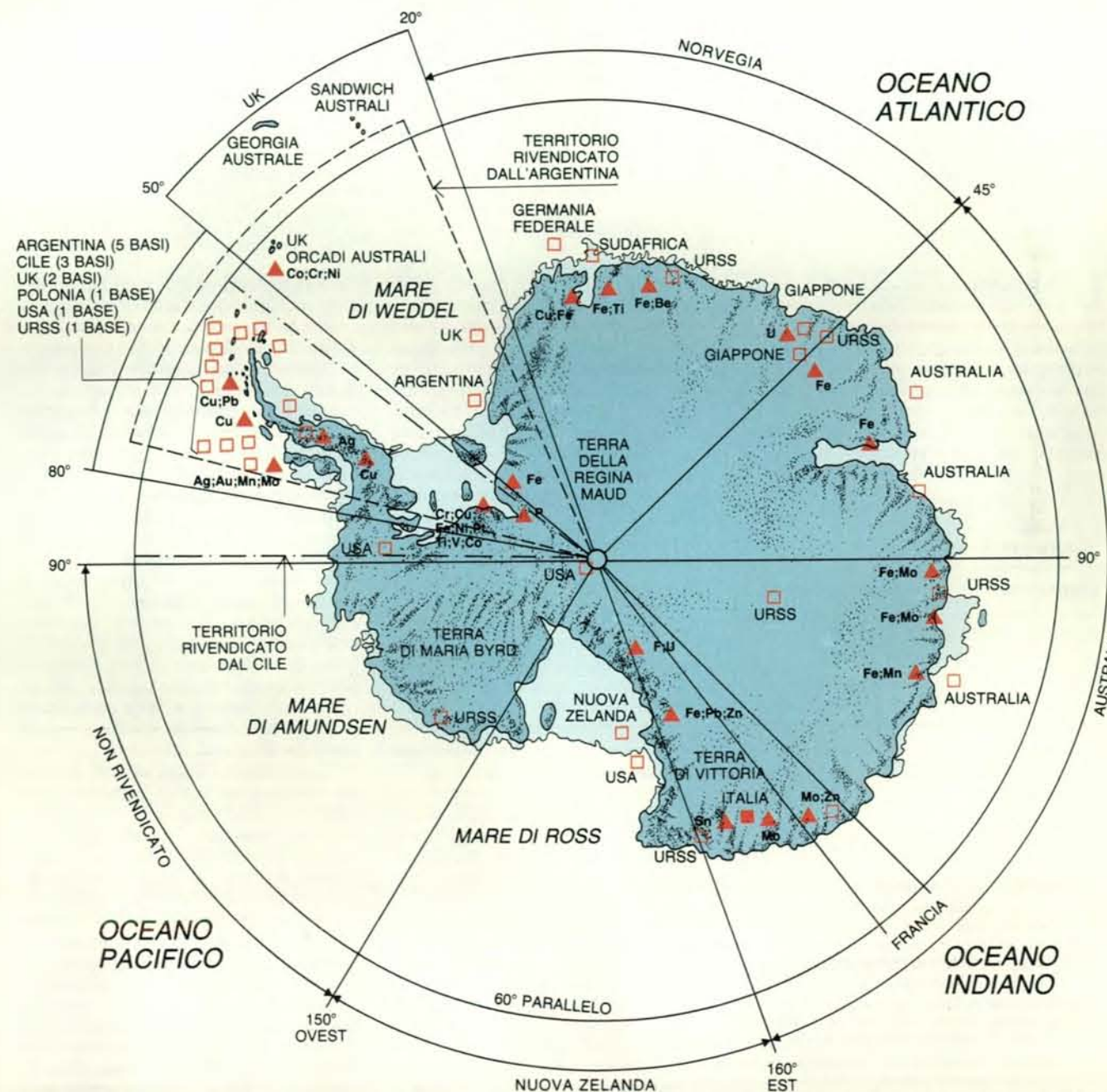
È invece di fondamentale importanza tenere conto di tali fattori qualora si voglia procedere a una corretta riflessione sull'evoluzione futura del settore minerario e metallurgico in Europa. Un primo problema da affrontare è quello delle riserve e degli investimenti. La valorizzazione delle restanti risorse del sottosuolo comunitario è ovviamente da incoraggiare e promuovere con mezzi adeguati sebbene, data la carenza di materie prime e la forte dipendenza dalle importazioni sopraccennata, sembri ragionevole aspettarsi che tale valorizzazione non possa offrire comunque prospettive sostanziali e di lungo periodo alle economie europee nel loro complesso. In merito alle riserve, è noto che la loro valutazione varia sensibilmente a seconda della materia prima considerata e a seconda di criteri quali la sfruttabilità economica, l'andamento dei prezzi, i progressi nelle tecniche di esplorazione e di coltivazione dei giacimenti, il prezzo del minerale utile (che cambia in modo fondamentale al variare dei costi di produzione). Pertanto, prima di affrontare un qualsiasi eventuale programma di valorizzazione dei giacimenti esistenti nei paesi della Comunità, occorre prevedere l'andamento di sviluppo dei fabbisogni di ciascun minerale e metallo considerato, che è a sua volta variabile a seconda dell'andamento della domanda globale e della produzione industriale.

Occorre inoltre considerare che non si dispone ancora di un inventario completo delle risorse e, quindi, delle possibilità di approvvigionamento interno delle economie europee in questo settore. A tal fine, occorrerebbe procedere a una ricerca mineraria di base, possibilmente privilegiando le aree economicamente meno sviluppate della Comunità (quali il Mezzogiorno d'Italia, l'Irlanda e la Grecia) e quelle carenti di tradizioni minerarie (Italia, Portogallo), utilizzando tecniche moderne e relativamente poco costose di ricerca (come le prospezioni geochimiche e geofisiche) e potenziando lo sviluppo della geologia e della scienza dei giacimenti.

A proposito delle ricerche minerarie, sono note le difficoltà incontrate in questi ultimi anni dagli investimenti privati: infatti, quelli minerari sono investimenti ad alto rischio, elevato immobilizzo di capitale e redditività molto dilazionata nel tempo. Pertanto in questi ultimi anni si è registrata quasi ovunque la contrazione degli investimenti privati, segnatamente di quelli destinati ai paesi in via di sviluppo. L'intervento pubblico, e in particolare delle istituzioni comunitarie, sarebbe perciò di grande importanza in questo campo; da un canto, esso potrebbe favorire un programma comune che offra alle imprese garanzie concrete contro parte dei rischi collegati alla ricerca, all'esplorazione mineraria e all'eventuale sfruttamento di risorse sia nel territorio comunitario sia presso paesi terzi (si pensi alla eventuale partecipazione finanziaria di organismi quali la Banca europea degli investimenti); dall'altro, accordi ufficiali di cooperazione con i paesi produttori primari potrebbero incentivare gli investimenti.

Un esempio di intervento pubblico in questo campo è quello fornito dalla recente legge mineraria italiana, n. 752 del 6 ottobre 1982, che prevede un finanziamento diretto da parte dello Stato per la ricerca di base su tutto il territorio nazionale e devolve anche contributi statali fino al 70 per cento a ricerche minerarie operative, sia in Italia sia presso paesi terzi, nei casi in cui esistano titoli minerari. È ancora presto per conoscere le conseguenze pratiche di questa nuova normativa, poiché alla ritardata approvazione della legge 752/82 si sono aggiunti ritardi burocratici e applicativi; ma i risultati potrebbero essere positivi se i finanziamenti previsti verranno utilizzati adeguatamente e non, magari, per risanare i vecchi bilanci delle società minerarie di Stato.

METALLI	AUTOSUFFICIENZA PRIMARIA (PERCENTUALE)	AUTOSUFFICIENZA PRIMARIA+SECONDARIA (PERCENTUALE)	IMPORTAZIONI AL NETTO DELLE ESPORTAZIONI (PERCENTUALE)
ALLUMINIO	27	49	55
RAME	0	43	58
PIOMBO	9	55	40
STAGNO	8	37	70
ZINCO	25	56	48
FERRO	7	54	45
MANGANESE	1	1	98
COBALTO	0	11	17
CROMO	1	14	86
MOLIBDENO	0	9	92
NICHEL	4	25	75
VANADIO	0	4	94
TUNGSTENO	17	40	49
ANTIMONIO	4	36	62
TITANIO	0	1	99

La scarsità di materie prime minerarie che caratterizza la Comunità economica europea è illustrata in percentuale del grado di copertura dei consumi interni per quanto riguarda i 15 metalli principali per uso industriale. I valori della seconda colonna (autosufficienza primaria) sono dati dal rapporto tra produzione primaria e consumo interno; quelli della terza colonna (autosufficienza primaria più autosufficienza secondaria) dal rapporto tra la produzione primaria più quella secondaria (riciclaggio) e il consumo interno; nella quarta colonna i valori si riferiscono al rapporto tra le importazioni della Comunità al netto delle esportazioni e il consumo interno. Anche se la produzione secondaria interna migliora il grado di autosufficienza, la dipendenza dalle importazioni da paesi esterni alla Comunità è compresa, se si eccettua il cobalto, tra il 40 e il 99 per cento. (La fonte dei dati è *Bilanci delle materie prime, 1979-1982*, EUROSTAT.)

Nella cartina è indicata la suddivisione provvisoria dell'Antartide, così come risulta dal Trattato antartico negoziato nel 1959 a Washington, e l'ubicazione delle basi nazionali dei vari paesi (*quadratini*). La base italiana (*quadratino pieno*) si trova nella Terra di Vittoria. Con i triangolini, affiancati dai simboli degli elementi principali, sono indicate le più importanti segnalazioni relative a giacimenti minerari metalliferi.

METALLI	PRODUZIONE PRIMARIA		PRODUZIONE SECONDARIA		IMPORTAZIONI		CONSUMI	
	1978	1982	1978	1982	1978	1982	1978	1982
ALLUMINIO	571	1141	959	898	3876	3534	4584	4204
RAME	6	4	832	1125	2188	2163	2514	2611
PIOMBO	147	111	650	574	651	699	1163	1245
STAGNO	3	4	12	15	61	42	68	53
ZINCO	414	375	531	473	1080	1164	1780	1525
FERRO	12360	6893	56626	49559	79381	75012	117614	105407
MANGANESE	2	6	–	–	1440	1171	1310	1083
COBALTO	–	–	–	–	17	6	6	4
CROMO	–	6	50	71	519	488	530	534
MOLIBDENO	–	–	2	2	36	37	29	26
NICHEL	–	6	33	31	180	149	183	151
VANADIO	–	–	–	–	10	10	5	8
TUNGSTENO	1	1	1	1	6	5	7	5
ANTIMONIO	1	1	10	5	21	16	24	17
TITANIO	–	–	2	2	481	565	421	391

L'analisi dei dati (in migliaia di tonnellate) relativi alla produzione, all'importazione e al consumo dei principali metalli di impiego industriale nella CEE, negli anni 1978-1982, consente alcune considerazioni di tipo qualitativo per quanto riguarda la politica di approvvigionamento primario seguita dalla Comunità. Per esempio, nonostante che, a causa della crisi di sovraproduzione dell'industria siderurgica, nel 1982 la produzione primaria di ferro si sia ridotta grosso modo alla metà di quella del 1978, circa l'81 per cento delle importazioni totali è costituito ancora da minerali di ferro. Anche per lo zinco e l'alluminio, settori di mercato in cui si ha un eccesso di capacità produttiva, si registra un'analoga tendenza; si privilegia cioè l'importazione di prodotti primari da indirizzare alla trasformazione di base senza tenere nel dovuto conto alcuni fattori significativi nel medio-lungo periodo, come la concorrenza internazionale, l'industrializzazione dei paesi emergenti, la situazione economica e di mercato dell'industria di trasformazione. (I dati sono tratti da *Bilanci delle materie prime 1979-1982*, EUROSTAT.)

Una promettente futura fonte di approvvigionamento primario è costituita dallo sfruttamento delle risorse minerarie dei fondali marini. Come è noto, la presenza di noduli polimetallici nei fondali oceanici fu segnalata per la prima volta dalla crociera scientifica compiuta dalla nave oceanografica inglese *Challenger* nel 1873, ma l'attenzione maggiore su di essi è stata richiamata solo dopo la metà del nostro secolo, proprio a causa della già accennata crisi della produzione mondiale di materie prime. Questi noduli, che si rinvengono nelle piane abissali oceaniche, constano in generale di una mescolanza di cristalli di vari minerali che inglobano talvolta materiali detritici. La loro struttura è sovente concentrica: spesso vi si può riconoscere un nucleo costituito da pomice, vetro vulcanico, frammenti litoidi di varia natura o resti scheletrici di organismi. I costituenti metallici sono in prevalenza manganese, ferro (circa il 25 per cento) e metalli pregiati quali nichel, rame, cobalto.

COMPONENTI	PERCENTUALE
ACQUA DI IDRATAZIONE	33
METALLI PREGIATI:	
NICHEL	0,98
RAME	0,87
COBALTO	0,15
ALTRI METALLI (MANGANESE, FERRO, ALLUMINIO ECC.)	23
SCORIE (SILICE, ARGILLE ECC.)	42

A partire dagli anni cinquanta, con l'accentuarsi della crisi che ha investito la produzione mondiale di materie prime, molta attenzione è stata dedicata ai cosiddetti noduli polimetallici, depositi metalliferi di forma arrotondata che si accumulano sui fondi oceanici attraverso processi diversi. Mentre sono in corso molte ricerche ed esplorazioni per valutare l'entità di questi depositi, per individuarne la distribuzione e per studiare la convenienza di un loro eventuale sfruttamento, poco si sta facendo a livello internazionale per disciplinare politicamente e giuridicamente l'accesso a questo tipo di risorsa. La tabella riporta la composizione media di un tipico nodulo polimetallico.

Compagnie minerarie europee, statunitensi e canadesi hanno da tempo intrapreso attività di esplorazione e ricerca nei fondali oceanici, mentre manca ancora del tutto un inquadramento di tipo politico ed economico sovranazionale: da anni infatti la conferenza dell'ONU sul diritto del mare procede stentatamente senza ottenere risultati operativi. La Comunità europea dovrebbe essere attivamente presente in questo campo: gli organi comunitari potrebbero contribuire al finanziamento di progetti, ma soprattutto a disciplinare in ambito sovranazionale lo sfruttamento delle risorse dei fondali marini mediante una migliore valutazione della più opportuna utilizzazione dei noduli (per esempio quali sostanze privilegiare rispetto ad altre, tenendo conto delle ricadute ecologiche).

Considerazioni analoghe potrebbero farsi per il futuro eventuale sfruttamento delle risorse minerarie presenti in Antartide. Da tempo vari paesi industrializzati e non (USA, URSS, Gran Bretagna, Francia, Cile, Argentina, Australia, Norvegia e Polonia) hanno iniziato indagini scientifiche su questo continente poco conosciuto, negoziando nel dicembre 1959 a Washington il «Trattato antartico». Questo prevede innanzitutto di accantonare i problemi concernenti rivendicazioni territoriali e assetto politico del continente, escludendo anche qualunque intervento riguardante l'utilizzazione militare di quel territorio, e prevede altresì un rinegoziato nel 1991 da parte di tutti i paesi che entro quella data avranno stabilito una base fissa di studi ed esplorazioni nel continente. Dei paesi che finora hanno ottemperato a queste condizioni fa parte anche l'Italia, che ha aderito al Trattato nel 1980 e successivamente ha installato una base nella Terra di Vittoria avviando un piano di ricerche geofisiche, geologiche, biologiche e geominerarie in vista anche della prospettiva di rinvenire risorse energetiche e metallifere di interesse.

La Comunità è finora assente da questa iniziativa, anche se il Parlamento europeo ha sollecitato la Commissione esecutiva ad assumere «un atteggiamento comune da parte degli Stati membri per quanto riguarda il futuro sfruttamento economico dell'Antartide». Per questo motivo sarebbe oltremodo auspicabile che prima del 1991 la Comunità affrontasse il problema accordandosi con i paesi membri già presenti in Antartide al fine di delineare un progetto comunitario di studi e ricerche oppure di creare un'«impresa comune» come previsto dal trattato CEEA o dall'Atto Unico del Lussemburgo (1986). Sarebbe certamente un grande successo per la Comunità se essa potesse sostituirsi direttamente a quei paesi membri già operanti nel continente antartico per svilupparvi un vero programma europeo.

Una fonte di approvvigionamento, da alcuni anni in costante espansione, è costituita dalla produzione secondaria di minerali e metalli, cioè dal riciclaggio di rottami e rifiuti di lavorazione, grazie anche a nuove tecniche che hanno avuto di recente un notevole sviluppo. Il ricorso al riciclaggio presenta alcuni ovvi e notevoli vantaggi: innanzitutto, diminuisce gli sprechi e aumenta sensibilmente la disponibilità di materia prima; inoltre, consente un notevole risparmio di energia rispetto al materiale non lavorato.

A titolo di esempio, basti considerare il caso della metallurgia dell'alluminio che è quella che richiede i più elevati costi energetici: l'energia difatti incide per oltre il 30 per cento sul costo totale del metallo. L'alluminio prodotto dal rottame, invece, consuma tra il 5 e il 10 per cento dell'energia necessaria alla produzione a partire dal minerale. Negli Stati Uniti, tra il 1970 e il 1983 la produzione secondaria di alluminio è cresciuta dal 20 al 33 per cento della produzione totale, ed è basata principalmente sul riciclaggio di recipienti in alluminio. Inoltre, il riciclaggio permette un risparmio sui costi di trasporto del minerale talora da aree lontanissime: si pensi, per esempio, all'industria primaria dell'alluminio in Italia, che importa la bauxite dagli antipodi (Australia), mentre il paese è tributario all'estero per oltre l'80 per cento dell'energia di cui ha bisogno.

Il riciclaggio infine costituisce anche un mezzo di salvaguardia ecologica, in quanto consente di evitare inquinamenti delle acque e del suolo a opera di taluni metalli nocivi, quali cromo, mercurio, piombo, sovente in parte perduti o indiscriminatamente scaricati nelle acque sotterranee durante le lavorazioni. In proposito, la citata proposta della Commissione CEE al Consiglio del luglio 1985 dedica molta attenzione al potenziamento delle tecnologie di riciclaggio, segnatamente ai fini ecologici, mediante investimenti e sforzi comuni.

Come si è accennato di sopra, una delle caratteristiche dell'evoluzione del settore metallurgico è costituita dai mutamenti in corso nella composizione relativa della domanda di minerali e metalli per uso industriale, dovuti allo sviluppo di nuove tecnologie e all'ingresso sul mercato di materiali avanzati. Si tratta fondamentalmente di un processo di sostituzione che è a sua volta un processo di risparmio energetico, dato che i nuovi materiali permettono una diversa e migliore utilizzazione delle risorse energetiche, aumentandone il rendimento (si veda in proposito il fascicolo di «Le Scienze» n. 220, dicembre 1986, interamente dedicato ai nuovi materiali per lo sviluppo economico). Basti pensare agli acciai speciali, alle superleghe e alle materie plastiche utilizzate nella costruzione delle automobili, che, consentendo una notevole riduzione del peso del veicolo, riducono il consumo di combustibile. Anche i nuovi materiali leggeri impiegati nell'industria aerospaziale, per esempio le speciali leghe di alluminio, nichel e titanio, oppure i materiali compositi costituiti da fibre di grafite in matrici polimeriche, consentono un risparmio di combustibile di circa il 30 per cento. Lo sviluppo di questi materiali avanzati è stato reso possibile da nuove tecniche di produzione e di trattamento dei metalli (come, per esempio, la metallurgia delle polveri, la solidificazione rapida, la solidificazione direzionale). Questa evoluzione tecnologica si sta già riflettendo sull'uso relativo di certi minerali e metalli, e di materiali in genere, rispetto ad altri. Un caso esemplificativo è quello del rame che sta venendo progressivamente sostituito dalle fibre ottiche come materiale di costruzione dei cavi per telecomunicazioni. Si prevede infatti che negli Stati Uniti, a causa dell'introduzione delle fibre ottiche, nei prossimi dieci anni il consumo di rame si ridurrà della metà.

Questi esempi confermano la necessità di tenere conto di vari fattori collegati all'evoluzione del settore minerario se si vuole tentare di programmare con strategie di lungo periodo non solo la produzione primaria, ma anche la politica industriale.

Il problema di fondo con il quale occorre confrontarsi nel trattare le prospettive del settore minerario e metallurgico è appunto un problema di politica industriale, che riguarda il futuro dell'industria di trasformazione europea. Infatti, la grave crisi che attraversano la siderurgia e la metallurgia dei metalli non ferrosi in tutti i paesi europei pone un problema di ristrutturazione e di razionalizzazione che costituisce uno degli imperativi industriali dei prossimi anni, come dimostrano i processi già in corso negli altri paesi industrializzati, Stati Uniti (assai più ricchi di risorse primarie ed energetiche dell'Europa), Giappone, Canada o i paesi scandinavi.

Il caso del ferro è certamente quello più evidente: la produzione di ferro e l'industria siderurgica europea attraversano da tempo una crisi di enormi proporzioni che né il ricorso a soluzioni protezionistiche né iniziative anche comunitarie di sostegno (come, per esempio, il cosiddetto «piano Davignon») sono riusciti ad arginare, dato che la siderurgia è certamente il settore più colpito dalla concorrenza internazionale, e in particolare dalle produzioni provenienti dai paesi in via di industrializzazione. L'occupazione nelle miniere europee di ferro si è ridotta nel 1983 a un quinto del valore del 1974, mentre nello stesso decennio gli addetti alla siderurgia sono diminuiti di oltre il 40 per cento (in Giappone del 20 per cento e negli Stati Uniti di oltre il 50 per cento). Soltanto in Italia l'occupazione si è ridotta meno che altrove a causa dei massicci investimenti (Bagnoli, Gioia Tauro) effettuati nel corso degli anni settanta, che oggi però fanno sentire tutti i loro effetti negativi e costituiscono un grosso ostacolo all'ammodernamento del settore. D'altra parte, per il caso italiano, basta scorrere le cifre delle perdite e degli indebitamenti di alcuni comparti delle partecipazioni statali per rendersi conto dell'entità del problema e degli effetti deleteri dell'assenza di una strategia di moderna politica industriale. Nell'ambito delle industrie statali, la siderurgia dell'IRI (gruppo Finsider) è quella che presenta infatti i peggiori risultati economici, avendo accusato perdite superiori ai 5000 miliardi

Nella CEE, il grado di occupazione nelle miniere di ferro e nell'industria siderurgica ha subito sensibili variazioni dal 1974 al 1983, come si rileva dall'istogramma. I rettangoli a sinistra si riferiscono al totale dei dipendenti nella siderurgia nei vari anni; quelli a destra al totale degli occupati nelle miniere di ferro. (I dati sono tratti da *Annuario siderurgia 1984*, EUROSTAT.)

MATERIALI	TENDENZA DI IMPIEGO		
	INCREMENTO		DIMINUZIONE
ACCIAI			DEBOLE
ACCIAI DOLCI		STABILE	
ACCIAI SPECIALI		STABILE	
ALLUMINIO E LEGHE	FORTE		
CERAMICHE TECNICHE	FORTE		
FIBRE	MOLTO FORTE		
GHISA		STABILE	
MAGNESIO E LEGHE	DEBOLE		
MATERIALI COMPOSITI	MOLTO FORTE		
METALLI PREZIOSI	FORTE		
NICHEL			DEBOLE
PIOMBO			DEBOLE
PLASTICHE	FORTE		
RAME		STABILE	
STAGNO			FORTE
TERRE RARE	DEBOLE		
TITANIO E LEGHE	FORTE		
ZINCO		STABILE	

Nell'attuale fase di sviluppo dell'impiego di materie prime, le applicazioni tecnologiche stimolano la domanda di nuovi materiali e la richiesta di prestazioni più avanzate da parte dei materiali tradizionali. In tabella sono riportate le tendenze di impiego al 2000 di alcuni materiali.

di lire nel solo triennio 1980-1982, a cui vanno aggiunte le migliaia di miliardi di sovvenzioni CEE.

Le cifre di altre produzioni metallurgiche sono altrettanto rivelatrici: il solo comparto dell'alluminio ha assorbito un terzo (oltre 1000 miliardi) dell'indebitamento finanziario dell'EFIM nel 1983, chiudendo il bilancio del biennio 1981-1982 con oltre 500 miliardi di perdite, senza contare le tariffe agevolate dell'elettricità, imposte dal Governo all'ENEL per questi impianti.

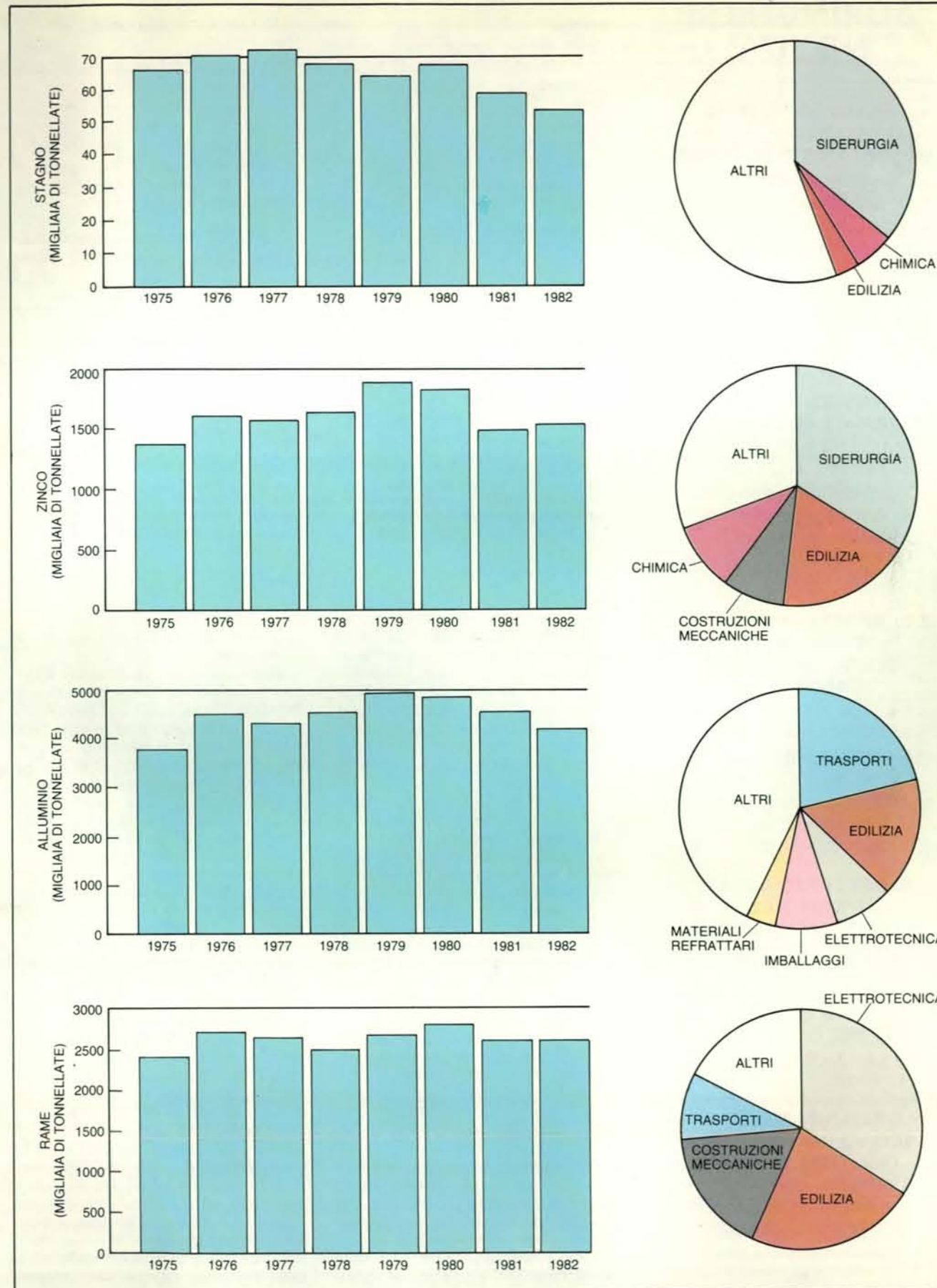
Il caso dell'alluminio è per certi versi altrettanto emblematico di quello della siderugia, trattandosi di una metallurgia fortemente energivora e inquinante. Eppure, come si è detto, la bauxite e altri minerali di alluminio costituiscono ancora il 47 per cento delle importazioni totali di alluminio della Comunità europea. Il 43 per cento della bauxite importata in Europa proviene dall'Australia e il 30 per cento dalla Guinea, con la conseguente notevole incidenza dei costi di trasporto sui costi finali. Trattandosi di un'industria ad alta intensità di capitale, marcatamente oligopolistica e soggetta a forti economie di scala, i paesi produttori primari hanno finora incontrato notevoli difficoltà nell'avviare e gestire i processi di prima trasformazione. Tuttavia, dati gli altissimi costi energetici, si prevede che in futuro l'energia inciderà sempre di più sulla localizzazione degli impianti, secondo una tendenza già oggi in atto: attualmente, il 58 per cento dell'alluminio prodotto nei paesi in via di sviluppo proviene da paesi non produttori di bauxite, ma che godono di un elevato potenziale energetico, particolarmente di origine idroelettrica. L'alluminio medio-orientale e canadese ha un costo inferiore del 15 per cento a quello europeo. Il Giappone è stato il primo paese a delocalizzare i suoi impianti di raffinazione mediante accordi di *joint venture* con i paesi in via di sviluppo del Sud-Est asiatico.

La Comunità europea è anche del tutto priva di rame, a parte la produzione spagnola. Stati Uniti, Giappone e Canada hanno da tempo avviato un'opera di razionalizzazione dell'industria di prima trasformazione, incrementando i procedimenti di seconda fusione. Quello del rame è uno dei settori nei quali migliori appaiono le prospettive di trasformazione presso i produttori primari. Infatti, l'industria del rame richiede costi di investimento inferiori a quelli di trasformazione di altri minerali ed è relativamente meno concentrata. I costi di trasporto incidono ovviamente molto meno per il rame raffinato e per i prodotti di prima lavorazione che non per i concentrati di miniera. Tra i maggiori produttori mondiali di rame figurano paesi (come Cile, Zambia, Zaire, Perù, che sono anche i principali fornitori della Comunità europea) con i quali sarebbe più che mai opportuno consolidare rapporti di cooperazione per la trasformazione primaria in loco con ovvi vantaggi reciproci, sempre che non ostassero ragioni politiche.

Considerazioni analoghe valgono per il piombo e lo zinco, un settore nel quale in Italia negli ultimi 10 anni il numero degli addetti si è ridotto del 62 per cento. Il piombo è molto inquinante e tende sempre più a essere sostituito da materiali più leggeri e meno inquinanti. Negli Stati Uniti, negli ultimi anni, si sono concentrati gli investimenti per potenziare la produzione di piombo di seconda fusione. Per quanto riguarda la Comunità europea, invece, il 40 per cento delle importazioni di piombo è ancora costituito da minerale, proveniente soprattutto da Canada, Sud Africa e Perù. Per quanto riguarda lo zinco, il 75 per cento delle importazioni comunitarie è costituito da minerale di zinco, proveniente in prevalenza da Australia e Perù. In genere le barriere protezionistiche allo zinco raffinato sono elevate, mentre alcuni paesi produttori primari, come Messico, Perù e Marocco, dispongono di capacità di raffinazione inutilizzate.

La sintetica serie di osservazioni che precedono dovrebbe rappresentare il punto di partenza dal quale avviare una riflessione sulle prospettive reali dell'approvvigionamento primario in minerali e metalli per le economie europee. Tali prospettive, come si è detto, vanno di pari passo con la più generale evoluzione tecnologica e l'industrializzazione dei paesi in via di sviluppo.

PAESI	ALLUMINIO RICICLATO (PERCENTUALE)					
	1954	1960	1965	1970	1975	1981
FRANCIA	21	19	17	20	23	27
GERMANIA FEDERALE	45	30	39	34	30	33
ITALIA	27	29	33	37	41	41
REGNO UNITO	30	32	36	35	30	28
GIAPPONE	19	26	29	28	–	25
STATI UNITI	18	16	17	17	21	28
STIMA MONDIALE	19	16	17	17	21	28

L'andamento del riciclaggio dell'alluminio è correlato a quello dei costi energetici, come si può constatare dai dati relativi ad alcuni paesi europei e agli Stati Uniti. Fa eccezione il Giappone per vari motivi: in quel paese il consumo di alluminio pro capite è stato basso fino all'ultimo decennio, quasi tutti i contenitori per bevande sono in acciaio o in vetro, la maggior parte del consumo di alluminio riguarda l'industria di prodotti destinati in gran parte all'esportazione, come le automobili. I paesi elencati in tabella rappresentano il 50 per cento della produzione mondiale di alluminio. I tassi di riciclaggio, dati come percentuale del consumo totale, si riferiscono al recupero secondario, compresi i rottami di alluminio impiegati sia nella produzione primaria sia in quella secondaria e le esportazioni nette di rottami. (I dati sono tratti da *Materials Recycling: the Virtue of Necessity* di William U. Chandler, Worldwatch Institute, Washington, 1983.)

L'andamento del consumo di alcuni metalli per uso industriale (stagno, zinco, alluminio e rame) nei paesi della CEE dal 1975 al 1982 è indicato in migliaia di tonnellate negli istogrammi a sinistra, mentre a destra sono illustrati, relativamente al 1981, i rispettivi settori di impiego.

METALLI	CONSUMI PRO CAPITE (CHILOGRAMMI)			
	1975	1978	1980	1982
ALLUMINIO	15	18	19	15
RAME	9	10	11	10
PIOMBO	4	4	5	5
ZINCO	5	7	7	6
FERRO	494	454	456	388
MANGANESE	6	5	5	4
CROMO	2	2	3	2
TITANIO	1	2	1	1
	CONSUMI PRO CAPITE (GRAMMI)			
	1975	1978	1980	1982
STAGNO	257	264	262	197
COBALTO	17*	24	24	17
MOLIBDENO	97	114	103	99
NICHEL	687	706	871	556
VANADIO	26	22	39	33
TUNGSTENO	25	26	22	19
ANTIMONIO	135*	94	70	66

In alto sono riportati in chilogrammi i consumi pro capite nella CEE di alcuni metalli di largo impiego industriale; in basso, sono riportati in grammi i consumi pro capite, sempre nella CEE, di altri metalli impiegati nell'industria in quantitativi più ridotti. I valori con l'asterisco si riferiscono al 1976. (I dati sono tratti da *Bilanci delle materie prime 1975-1982*, EUROSTAT.)

L'Europa è molto indietro nei processi di ristrutturazione e razionalizzazione del settore; ma, soprattutto, sembra ancora molto indietro e molto esitante riguardo alle scelte di fondo di politica industriale e di programmazione energetica che dovrebbero essere indirizzate a migliorare la competitività delle produzioni europee. Tutto ciò sarà possibile soltanto puntando sulle lavorazioni più avanzate, ad alta intensità di capitale, che privilegino la ricerca tecnologica e la messa a punto di nuovi materiali. La vecchia tradizionale metallurgia di base, energivora e inquinante, è un'industria della quale occorre gradatamente sbarazzarsi, come stanno già facendo molti paesi industrialmente più progrediti (per esempio, Giappone e Svezia) per concentrare invece gli sforzi finanziari, tecnologici e di ricerca nelle fasi più avanzate dei cicli di lavorazione e nelle moderne tecniche di riciclaggio.

La delocalizzazione di certa metallurgia di base costituisce perciò una tendenza inevitabile nel lungo periodo, poiché è legata ai mutamenti del mercato internazionale e all'evoluzione strutturale del settore. Per queste ragioni i paesi europei e la Comunità nel suo complesso, invece di varare gli ennesimi piani di sovvenzioni, elevare barriere protezionistiche e far gravare sulla collettività i costi di un anacronistico uso dell'energia, dovrebbero promuovere politiche di cooperazione internazionale incoraggiando la trasformazione di base presso i paesi produttori primari, particolamente quelli che dispongono di risorse energetiche a basso costo (soprattutto idroelettriche e geotermiche), anche garantendo loro, nel contempo, l'esportazione di semilavorati.

Quanto detto non vuole essere un invito a rinunciare completamente alla metallurgia primaria in Europa, viste le tipiche caratteristiche di integrazione verticale di questa industria, in parte da salvaguardare per consentire un ulteriore sviluppo tecnologico e il mantenimento di una professionalità avanzata, e anche in considerazione degli effetti sociali delle attuali difficoltà occupazionali, che ovviamente non consentono bruschi cambiamenti di direzione. Tuttavia, il problema dell'approvvigionamento primario nel suo complesso e, di conseguenza, quello dell'industria di trasformazione di base, sono per eccellenza problemi di strategia industriale di lungo periodo così come lo è il problema energetico, al quale è strettamente collegato. Trattandosi di problemi strutturali della economia europea, essi vanno affrontati nella loro globalità, con la lungimiranza, l'equilibrio e la fantasia che sono sempre necessari per gestire al meglio i processi di sviluppo. Ma un progresso in questo senso richiede in primo luogo che i 12 paesi della Comunità europea rinuncino a ricercare soluzioni particolari e a sfondo talora nazionalistico, per perseguire anche in questo settore, come invano si reclama da tempo per l'energia, una politica comune che incida in profondità e vada ben più in là delle retoriche invocazioni all'unità europea che abbiamo ascoltato più volte negli ultimi anni e ancora di recente nel trentennale della firma dei Trattati di Roma.

I satelliti di Urano

Le immagini trasmesse dalla sonda Voyager 2 *ci hanno mostrato per la prima volta i cinque maggiori satelliti di Urano, rivelando che tre di questi corpi celesti hanno avuto in passato un'attività geologica molto intensa*

di Torrence V. Johnson, Robert Hamilton Brown e Laurence A. Soderblom

A un immaginario osservatore posto nel sistema di Urano, il passaggio di *Voyager 2* sarebbe apparso come un breve intermezzo, un evento relativamente insignificante nell'immutabilità del pianeta con i suoi anelli circolari e i suoi satelliti ghiacciati e coperti di crateri. Ma soprattutto a coloro che lo avevano lanciato, i tecnici del Jet Propulsion Laboratory (JPL) in California, il passaggio di *Voyager 2* il 24 gennaio 1986 è sembrato troppo breve. Questo è particolarmente vero per coloro che si occupavano dei cinque principali satelliti del pianeta, i quali sono appena visibili da Terra anche facendo uso dei telescopi più potenti. Quando *Voyager 2* è sfrecciato vicino a Urano a una velocità di 72 000 chilometri all'ora, vi era meno di un giorno di tempo per raccogliere quelle che in pratica rimarranno forse le informazioni più dettagliate che potremo mai avere sui suoi satelliti: al momento infatti non è prevista alcun'altra missione verso Urano.

Gli ostacoli tecnici erano formidabili; concepito in origine per operare solo fino alla distanza di Saturno, dove la luce del Sole è quattro volte più intensa e dove i satelliti ne riflettono in media una quantità doppia, *Voyager 2* aveva a che fare nel caso di Urano con oggetti molto più deboli di quelli per i quali le sue telecamere erano state progettate. Per ricevere la quantità di luce sufficiente per formare un'immagine, gli otturatori delle macchine da ripresa dovevano restare aperti per diversi secondi. A meno che non si fosse riusciti a trovare un qualche sistema per compensare l'effetto del moto della sonda e dei satelliti stessi, le immagini sarebbero risultate «mosse» al limite dell'inutilità. Per avere un'idea del problema al quale dovevano far fronte *Voyager 2* e coloro che lo dirigevano, si immagini di voler fotografare una macchina da corsa grigio scuro in una giornata buia e nuvolosa, facendo uso di una pellicola poco sensibile. Tuttavia il problema è stato completamente risolto. Mentre la sonda procedeva da Saturno verso Urano, i tecnici del JPL hanno escogitato un metodo per compensare i vari e complessi movimenti (si veda l'articolo *I problemi tecnici del volo di* Voyager 2 *verso Urano* di Richard P. Laeser, William I. McLaughlin e Donna M. Wolff in «Le Scienze» n. 221, gennaio 1987). I frutti di questi sforzi sono state immagini nitide e di alta qualità di tutti e cinque i principali satelliti di Urano: Oberon, Titania, Umbriel, Ariel e Miranda. Nel caso di Miranda, il più interno dei cinque, le immagini hanno raggiunto una risoluzione migliore rispetto a tutte quelle ottenute per Giove e Saturno da *Voyager 1* e da *Voyager 2*.

La brevità dell'incontro con Urano era una conseguenza non solo della velocità della sonda, ma anche della insolita disposizione nello spazio del sistema di Urano il cui asse di rotazione giace all'incirca sul piano dell'orbita. Attualmente, il polo sud del pianeta punta verso il Sole e la Terra. Poiché i satelliti orbitano all'incirca sul piano equatoriale, la traiettoria di *Voyager 2* attraverso il sistema era simile a quella di un proiettile che attraversa il bersaglio di un tiro a segno. Invece di passare vicino a ciascun satellite, uno dopo l'altro, come nel caso di Giove e Saturno, la sonda ha raggiunto la minima distanza da tutti i satelliti di Urano all'incirca nello stesso istante. Per assicurare una copertura di immagini fotografiche ad alta risoluzione di ogni satellite, la sequenza delle riprese ha dovuto essere calcolata con la precisione di un secondo. Anche con un adeguato programma operativo, tuttavia, non più di metà di ogni satellite ha potuto essere fotografata, dato che attualmente solo gli emisferi meridionali sono illuminati dal Sole.

Nei mesi che hanno preceduto il massimo avvicinamento della sonda, una grande quantità del tempo dedicato all'osservazione è stata impiegata nell'esplorazione del piano equatoriale del pianeta alla ricerca di nuovi satelliti. C'erano buoni motivi per sospettare la loro esistenza: Giove ha 16 satelliti, e Saturno 17, per la maggior parte oggetti minuscoli non visibili da Terra. *Voyager 2* ha scoperto 10 nuovi satelliti di Urano, portando il totale per questo pianeta a 15. Due dei nuovi satelliti sono «pastori» gravitazionali, che «pattugliano» i bordi esterno e interno dell'anello più grande e più esterno del pianeta. Gli altri otto percorrono orbite circolari comprese tra gli anelli e Miranda. Tranne uno, tutti i nuovi satelliti hanno diametri compresi tra 40 e 80 chilometri. L'unico che fa eccezione, indicato come 1985U1 in attesa dell'assegnazione di un nome ufficiale da parte dell'Unione astronomica internazionale, ha un diametro di 160 chilometri.

Poiché 1985U1 è stato scoperto più di un mese prima dell'incontro e poiché si trovava sullo stesso lato del pianeta in cui osservavamo Miranda, siamo riusciti, insieme ai nostri colleghi, a dedicargli parte del tempo-telecamera destinato in origine a Miranda, ricavando in tal modo un'immagine ravvicinata del nuovo satellite. Questa immagine mostra un oggetto piuttosto irregolare con una superficie coperta di crateri, tanto scura che riflette solo il 7 per cento della luce del Sole. Gli altri nuovi satelliti non sono stati osservati da vicino, ma sembrano essere ugualmente scuri, così come le particelle che compongono gli anelli del pianeta.

Da questo punto di vista, essi sono tutt'altro che peculiari. In realtà, la maggior parte del materiale solido nel sistema solare esterno che non sia ghiaccio brillante, è molto scuro, con un'albedo, o fattore di riflessione, di circa il 10 per cento o meno. La natura di questo materiale scuro è stata oggetto di discussione negli ultimi anni. La maggior parte degli studiosi è d'accordo sul fatto che questo materiale sia scuro perché è ricco di carbonio, come la fuliggine. La spiegazione più ovvia è che si tratti sostanzialmente dello stesso materiale che si trova nelle condriti carbonacee: meteoriti primordiali che provengono dalla fascia degli asteroidi tra Marte e Giove e consistono di silicati idrati, o argille, mescolati con sostanze organiche scure, ricche di carbonio. Tali miscele si formano solo a temperature relativamente basse. I modelli teorici indicano che le condizioni della nebulosa solare primordiale sarebbero state favorevoli alla formazione di rocce carbonacee a partire dalla fascia degli asteroidi verso l'esterno.

In alternativa, il materiale carbonaceo potrebbe non essere roccia primordiale, ma piuttosto il prodotto recente dell'irradiazione di ghiaccio di metano, o ghiaccio d'acqua contaminato con metano, che si ritiene possa formarsi nel sistema solare esterno. Studi di laboratorio hanno dimostrato che radiazioni ultraviolette o elettroni di alta energia (come quelli presenti nel vento solare o

Nella proiezione stereografica dell'emisfero meridionale di Miranda, ricavata dalle immagini trasmesse da *Voyager 2*, il polo sud è al centro e l'equatore è il bordo circolare esterno. La sonda ha potuto fotografare solo gli emisferi meridionali dei satelliti di Urano, dato che solo questi erano illuminati dal Sole. Miranda è il più piccolo e il più «strano» dei cinque satelliti principali. Sovrapposte alle sue antiche pianure craterizzate vi sono tre regioni più giovani, meno craterizzate, con terreni a solchi e creste, chiamate «regioni ovoidali». La superficie del satellite è anche attraversata da zone di frattura. Quella visibile nella parte alta dell'immagine comprende un dirupo alto tra 10 e 20 chilometri. Le proiezioni stereografiche che illustrano questo articolo sono state realizzate dall'US Geological Survey a Flagstaff sotto la direzione di R. M. Batson.

La traiettoria di *Voyager 2* era quasi perpendicolare al piano equatoriale di Urano. Poiché i satelliti orbitano presso il piano equatoriale, la sonda ha raggiunto la distanza minima da tutti i satelliti quasi nello stesso momento, per cui il tempo disponibile per riprendere immagini in primo piano è stato breve. Le linee in colore indicano le posizioni e i tempi (misurati in ore prima del massimo avvicinamento) in cui si sono ottenute le immagini migliori di ogni satellite.

nelle magnetosfere planetarie) possono provocare reazioni chimiche nelle quali il metano dà origine a un residuo scuro di composti organici complessi. Questo meccanismo è stato più volte invocato per spiegare la produzione di grani scuri interstellari, il basso fattore di riflessione dei nuclei delle comete e il peculiare lato scuro del satellite di Saturno, Giapeto. In effetti, alcuni studiosi pensano che questo meccanismo sia in grado di spiegare la presenza di materiale organico nelle rocce carbonacee. Se questo è vero allora l'interrogativo sulla formazione di composti scuri a partire dal metano nel sistema solare esterno non consiste più nello stabilire se essa ha avuto luogo, ma quando ciò è avvenuto: se è stata importante solo quattro miliardi e mezzo di anni fa, quando le rocce carbonacee si formarono, o se ha avuto luogo anche più tardi sulla superficie dei satelliti.

Finora non c'è alcuna particolare ragione per credere che questo processo abbia avuto un effetto significativo sulla superficie dei satelliti di Urano. Da un lato, la presenza di metano sui satelliti, a differenza del ghiaccio d'acqua, non è stata ancora provata con osservazioni da Terra. (*Voyager 2* non è equipaggiato con uno spettrometro in grado di analizzare la composizione chimica delle superfici dei satelliti.) A nostro parere, la maggior parte del materiale scuro nel sistema di Urano è probabilmente roccia carbonacea primordiale, anche se composti scuri derivati dal metano possono essersi aggiunti alla composizione originale in qualche luogo. Esperimenti eseguiti da Roger N. Clark, allora all'Università di Hawaii a Manoa, indicano che piccole quantità di roccia carbonacea (meno dell'uno per cento in volume), quando sono mescolate con ghiaccio d'acqua, sono sufficienti a rendere scuro il ghiaccio, portandolo a un fattore di riflessione dell'ordine di quel 7 per cento, mostrato da 1985U1.

I cinque maggiori satelliti di Urano non sono così scuri come 1985U1; le loro albedo vanno dal 20 per cento (Umbriel) al 40 per cento (Ariel). Questo fatto è significativo. Se si ammette che 1985U1 rappresenti la miscela primordiale di ghiaccio e roccia dalle quali si sono formati tutti i satelliti, si deve spiegare perché quelli più grandi siano più brillanti. La risposta è che 1985U1 è probabilmente una miscela omogenea di ghiaccio e roccia, mentre i satelliti più grandi non lo sono. Al contrario, essi hanno subito vari gradi di differenziazione: una parte della roccia densa è sprofondata verso il centro del satellite e una parte del ghiaccio, più leggero e più brillante, è risalito alla superficie. Il processo di differenziazione richiede del calore interno, e le sorgenti di calore in tutti i satelliti di Urano probabilmente divennero insignificanti molto tempo fa. Ma a giudicare dalle immagini di *Voyager 2*, tutti e cinque i satelliti - specialmente Ariel e Miranda - sembrano essere stati attivi dal punto di vista geologico agli albori della loro storia.

Le prove di attività geologica su un satellite ghiacciato possono assumere forme diverse. Se ha avuto luogo un'estesa differenziazione, si possono scorgere vaste configurazioni di zone chiare e scure sulla superficie in corrispondenza di regioni dove una minore o maggior quantità di roccia è mescolata al ghiaccio. Se la crosta del satellite è stata frantumata da movimenti tettonici, si dovrebbero scorgere delle linee di faglia. E se la sorgente di calore interno è stata abbastanza intensa da sviluppare un'attività vulcanica, si dovrebbero vedere regioni coperte da materiale relativamente giovane che è stato eruttato dall'interno del satellite ed è colato sulla superficie.

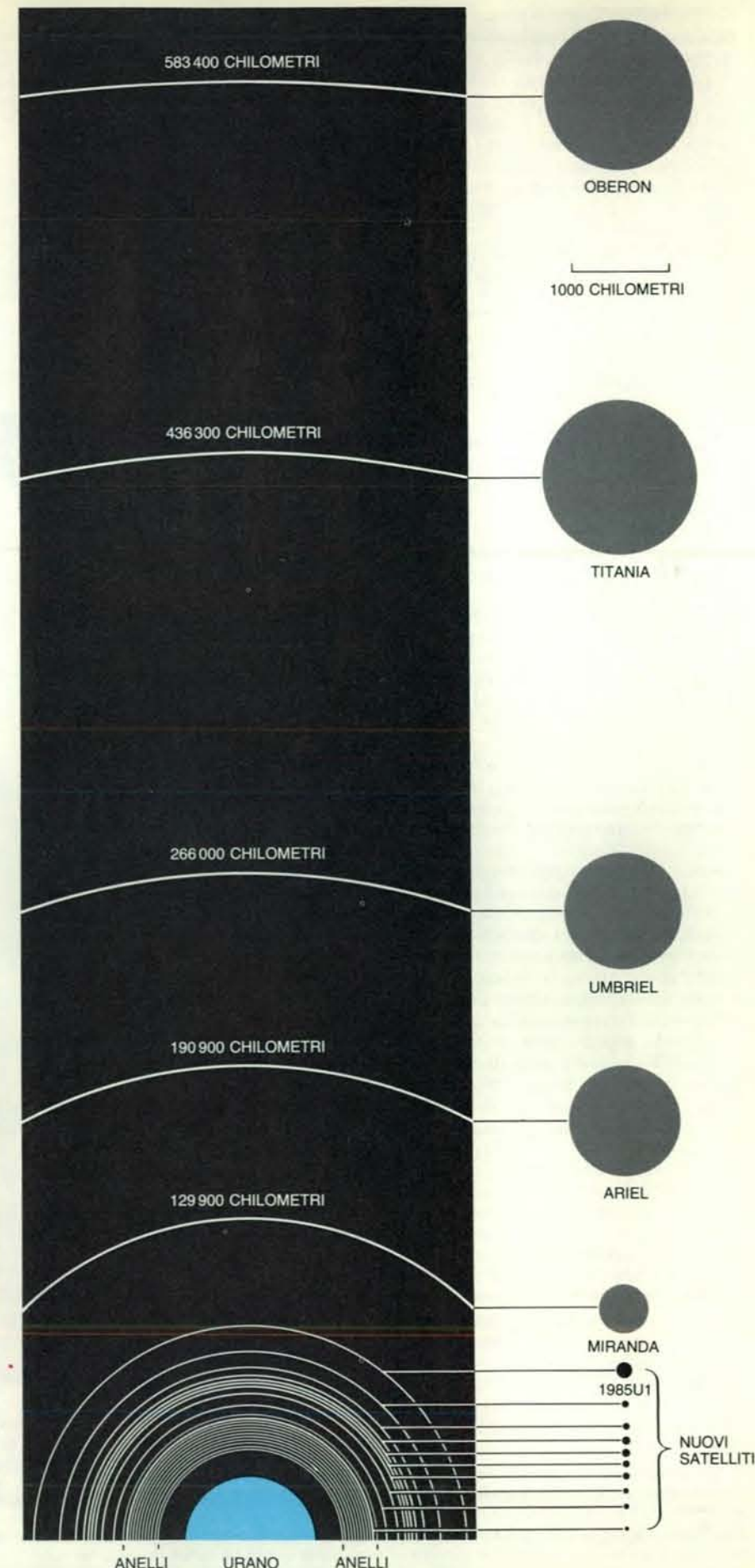
Che cosa si intende per «giovane»? Gli unici indicatori di tempo sulla superficie dei satelliti sono i crateri scavati dai massi di roccia e ghiaccio piovuti su di essi a partire dall'epoca in cui si formarono, quattro miliardi e mezzo di anni fa. Una regione del tutto coperta di crateri è ovviamente antica; una regione totalmente priva di crateri deve aver subito un ricoprimento superficiale in un passato geologicamente recente. Datare con maggiore precisione le caratteristiche geologiche, soprattutto se la loro età cade tra questi estremi, è difficile, dato che il tasso di craterizzazione è stato diverso da pianeta a pianeta e non è rimasto costante nel tempo. Il bombardamento più massiccio ebbe luogo quando i pianeti e i loro satelliti erano ancora masse in via di accrescimento.

Durante questo periodo iniziale, almeno due distinte popolazioni di proiettili, corrispondenti a due sorgenti di detriti, colpirono i satelliti nel sistema solare esterno. La prima popolazione era costituita da materiale che era rimasto in orbita intorno al Sole dopo che i pianeti si furono formati. I crateri scavati da questa popolazione hanno un'ampia distribuzione di grandezze; tipicamente, i loro diametri vanno da meno di un chilometro fino a centinaia di chilometri. Essi sono stati trovati in tutto il sistema solare nei terreni più antichi come gli altipiani della Luna. La seconda popolazione di oggetti consisteva di detriti rimasti in orbita intorno ai pianeti dopo la formazione dei loro satelliti. La maggior parte di questi corpi celesti era relativamente piccola. Prima dell'incontro con Urano, i crateri lasciati da questi proiettili erano stati scoperti principalmente sui satelliti di Saturno, soprattutto da Robert G. Strom e colleghi dell'Università dell'Arizona. I bombardamenti di oggetti delle due popolazioni si sono sovrapposti, ma quello della seconda popolazione sembra essere terminato un po' più tardi. Questo fatto è un utile indizio per la comprensione delle prime fasi della storia dei satelliti di Urano.

Oberon e Titania, i due satelliti più grandi e più esterni, hanno ambedue un diametro di circa 1600 chilometri (meno di metà di quello della Luna), e hanno all'incirca la stessa massa. Le loro superfici, ricche di ghiaccio d'acqua, riflettono tra il 25 e il 30 per cento della luce solare incidente e sono di colore grigio uniforme. Il grigio è interrotto solo da linee brillanti disposte a raggiera intorno ad alcuni crateri; nessuno dei due satelliti mostra configurazioni di zone chiare e scure su larga scala. Nonostante questi elementi di somiglianza, i due satelliti hanno avuto un'evoluzione chiaramente diversa.

Oberon possiede qua e là qualcosa di simile a faglie, ma non presenta che scarsi indizi di una significativa fratturazione di origine tettonica. La sua superficie è coperta quasi interamente di grandi, antichi crateri con dimensioni che vanno dal limite di risoluzione delle immagini (circa 12 chilometri), a più di 100 chilometri. Il fatto che i crateri grossi non siano stati cancellati indica che la superficie di Oberon non è stata ricoperta di materiale nuovo dai tempi del bombardamento di meteoriti della prima popolazione, che finì qualcosa come quattro miliardi di anni fa.

Questo non significa che il satellite sia

Il sistema di Urano comprende almeno 11 anelli e 15 satelliti. Le orbite sono misurate a partire dal centro di Urano, che ha un diametro di 25 600 chilometri; tutti i satelliti sono disegnati nella stessa scala. I cinque satelliti maggiori erano noti prima del viaggio di *Voyager 2*. Dei dieci piccoli satelliti scoperti in questa occasione, solo il più grande, 1985U1, è stato fotografato da vicino (*qui sopra*). La sua superficie è più scura di quella dei satelliti maggiori. Il grande cratere visibile sul bordo a destra ha un diametro di circa 40 chilometri.

stato completamente inattivo. Il fondo di alcuni dei crateri più vasti è coperto da chiazze di materiale molto scuro simile a quello che si osserva su 1985U1. Il loro aspetto ricorda quello dei crateri con fondo scuro, come Tsiolkovsky, sugli altipiani della Luna. Sulla Luna i depositi scuri sono stati probabilmente lasciati da eruzioni vulcaniche che ebbero luogo in seguito agli enormi impatti che hanno scavato i crateri. Anche Oberon potrebbe avere avuto una certa attività vulcanica primordiale localizzata. Le sue chiazze scure possono essere costituite da una miscela di ghiaccio e rocce carbonacee eruttate dall'interno. In alternativa, nelle colate vulcaniche era forse contenuto ghiaccio d'acqua contaminato da metano, che può essersi scurito dopo aver raggiunto la superficie.

Mentre Oberon è stato essenzialmente un bersaglio passivo per gli oggetti provenienti dall'esterno, questo non è il caso di Titania. La superficie di questo satellite porta segni molto evidenti di un'attività tettonica globale: un insieme complicato di fosse tettoniche (*graben*), delimitate da faglie di distensione. Inoltre, sebbene la superficie di Titania sia fortemente craterizzata, sembra che i crateri siano stati formati da corpi celesti della seconda popolazione. Titania deve avere subito un bombardamento da parte di oggetti della prima popolazione intenso almeno quanto quello di Oberon, ma oggi possiede solo un piccolo numero di grandi crateri che possono essere dovuti a questa popolazione. Evidentemente la maggior parte dei grandi crateri è stata cancellata da qualche processo di rimodellamento della superficie. Per di più, dato che la densità di crateri non è uniforme - ci sono diverse zone pianeggianti che sono chiaramente più lisce e meno craterizzate rispetto al resto della superficie - questo rimodellamento deve essere durato per un periodo di tempo considerevole.

Secondo uno dei modelli dell'evoluzione di Titania, il rimodellamento superficiale avvenne con una vasta estrusione vulcanica di materiale sulla superficie. Il vulcanismo iniziò mentre il bombardamento di meteoriti della prima popolazione era ancora in atto. Grandi crateri scomparvero perché vennero sommersi o semplicemente perché la crosta ghiacciata intorno a essi, che doveva essere ancora relativamente calda e malleabile, collassò. A mano a mano che il satellite irradiava il suo calore interno, esso iniziò a gelare a partire dalla crosta. Alla fine i liquidi, e per prima l'acqua, gelarono anche nell'interno. L'acqua si comportò come sempre quando gela a bassa pressione: cominciò ad aumentare di volume e, poiché era presente in grande quantità, la superficie del satellite subì stiramenti sempre più intensi. La crosta si ruppe lungo una rete di faglie di distensione e blocchi di crosta precipitarono lungo le faglie, formando le enormi fosse tettoniche. I movimenti tettonici possono essere stati accompagnati da ulteriori estrusioni di fluidi che probabilmente livellarono le pianure.

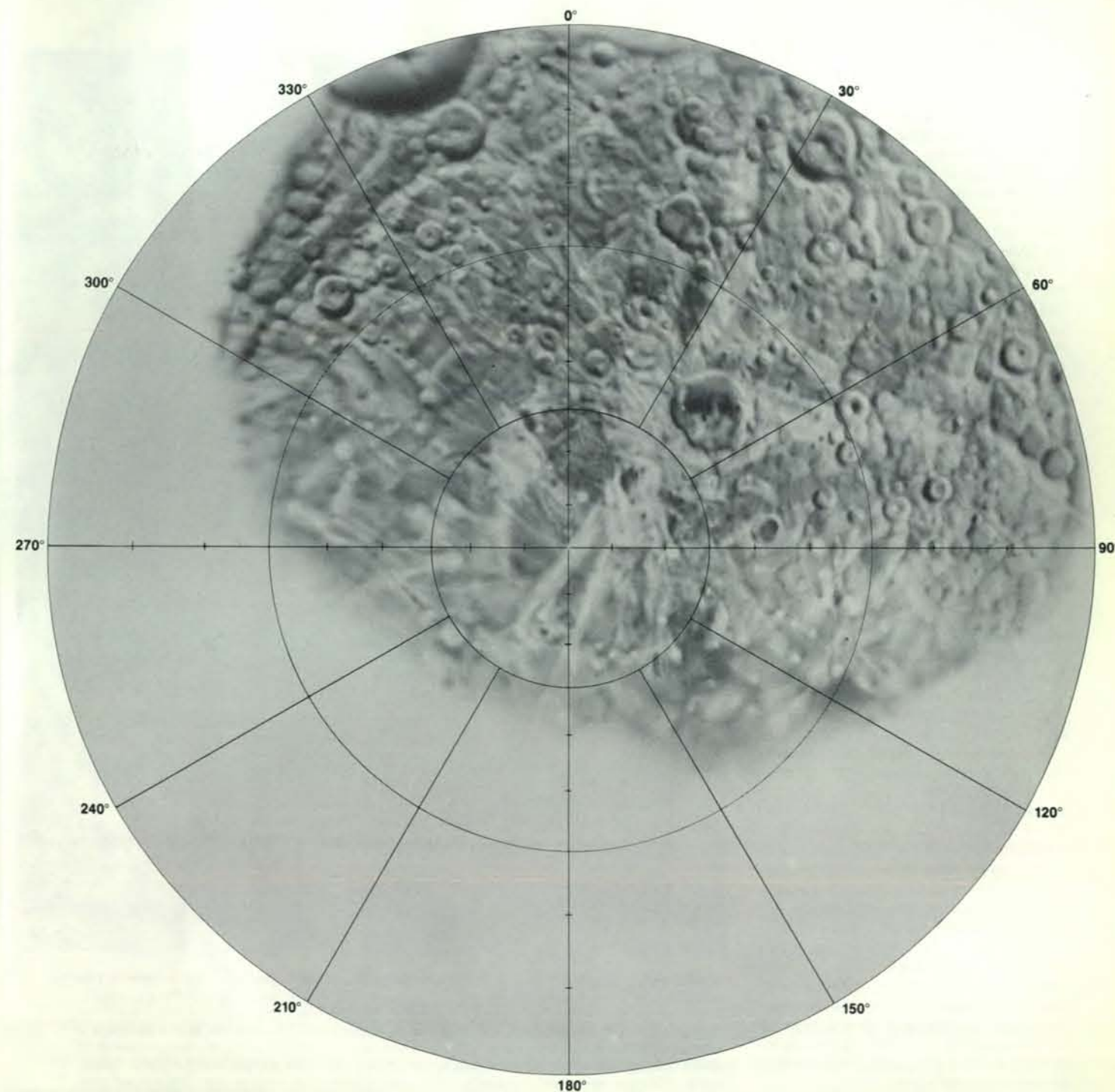
Quando la maggior parte della superficie ebbe terminato di rimodellarsi, il primo bombardamento era cessato e solo i corpi più piccoli della seconda popolazione continuarono a cadere su Titania. Alla fine, quando lo spazio intorno a Urano fu libero da detriti, anche questo bombardamento cessò. Nei successivi tre miliardi di anni all'incirca, Titania è stato un satellite tranquillo, disturbato solamente dall'urto di qualche sporadica cometa.

Eugene M. Shoemaker dell'US Geological Survey, un esperto di craterizzazione, ha proposto un tipo più drastico di rimodellamento della superficie. Egli ritiene che il primitivo processo di craterizzazione di Titania da parte di oggetti della prima popolazione sia stato tale che l'impatto di un grande corpo avvenuto verso la fine del processo potrebbe avere distrutto il satellite. I frammenti non si sarebbero allontanati dall'orbita di Titania e si sarebbero riaggregati relativamente in fretta in un nuovo satellite. Questo satellite riformatosi avrebbe avuto una nuova superficie che non portava traccia del bombardamento precedente. Un vantaggio dell'ipotesi di Shoemaker è che può spiegare perché Oberon e Titania abbiano proprietà complessive simili, ma il loro aspetto sia tanto diverso: essendo più vicino a Urano, il cui campo gravitazionale deve avere concentrato e accelerato i detriti in arrivo, Titania aveva una probabilità più alta di essere distrutto.

Umbriel e Ariel, i due successivi satelliti di Urano, costituiscono un'altra coppia interessante. Hanno quasi lo stesso diametro (1190 e 1160 chilometri rispettivamente) e la stessa massa, ma il contrasto tra loro è ancora maggiore di quello tra Oberon e Titania.

Umbriel è peculiare per la sua quasi totale uniformità. Come Oberon e Titania, non ha variazioni di luminosità su larga scala, ma, al contrario di questi due oggetti, la sua superficie non presenta crateri circondati da una raggiera brillante. (Si pensa che le strisce brillanti disposte a raggiera intorno ai crateri sui satelliti ghiacciati si formino perché gli impatti scagliano in superficie ghiaccio sotterraneo pulito.) Mentre *Voyager 2* si avvicinava al sistema di Urano, Umbriel sembrava del tutto privo di crateri, il che significa che la sua superficie dovrebbe tuttora continuare a essere rimaneggiata. Le immagini ad alta risoluzione riprese alla minima distanza, tuttavia, hanno mostrato una superficie dominata da vasti crateri derivanti dal bombardamento della prima popolazione di meteoriti; solo le raggiere sono assenti.

Spiegare l'uniformità di Umbriel è sorprendentemente difficile. Si può immaginare che le raggiere brillanti intorno ai suoi crateri abbiano potuto essere cancellate da micrometeoriti che bersagliarono la superficie rimescolando il

Oberon è stato fotografato da *Voyager 2* da una distanza di 660 000 chilometri. È solo di poco più piccolo di Titania, il maggiore dei satelliti di Urano, ma la sua antica superficie densamente craterizzata mostra scarsi segni di attività geologica. Depositi scuri, forse di origine vulcanica, sono visibili sul fondo di alcuni crateri, come quello grande vicino al centro della fotografia; nella proiezione stereografica dell'emisfero meridionale di Oberon (*a destra*), il cratere è situato tra 30 e 60 gradi di longitudine e a una latitudine di circa 50 gradi. Il cratere ha un diametro di più di 100 chilometri e, come i numerosi altri crateri di Oberon, risale al bombardamento da parte di detriti in orbita intorno al Sole, che terminò circa quattro miliardi di anni fa.

materiale delle raggiere con il materiale scuro sottostante. (Questo processo è chiamato *impact gardening*). Una seconda possibilità è che le raggiere contenessero metano che è diventato scuro per effetto di radiazioni energetiche. Ambedue queste ipotesi non riescono però a spiegare perché le raggiere brillanti siano assenti su Umbriel, ma non sugli altri satelliti di Urano.

Forse la spiegazione più semplice è che le raggiere brillanti non si sono mai formate su Umbriel. Se Umbriel fosse stato ricoperto da uno strato di materiale scuro spesso diversi chilometri, i frammenti originati a causa di un impatto sarebbero stati scuri; raggiere scure di questo tipo sono state osservate, per esempio, sul satellite di Giove, Ganimede. Il rivestimento potrebbe essere costituito da un'uniforme miscela primordiale di ghiaccio e roccia. Una copertura primordiale su Umbriel è per lo meno plausibile, dato che Umbriel è più piccolo di Oberon e Titania, e quindi ci si può aspettare che abbia subito una differenziazione minore. Per di più, il fatto che esso abbia il fattore di riflessione più basso (20 per cento) tra i satelliti di Urano favorisce l'ipotesi che abbia un'alta concentrazione di materiale carbonaceo scuro vicino alla superficie.

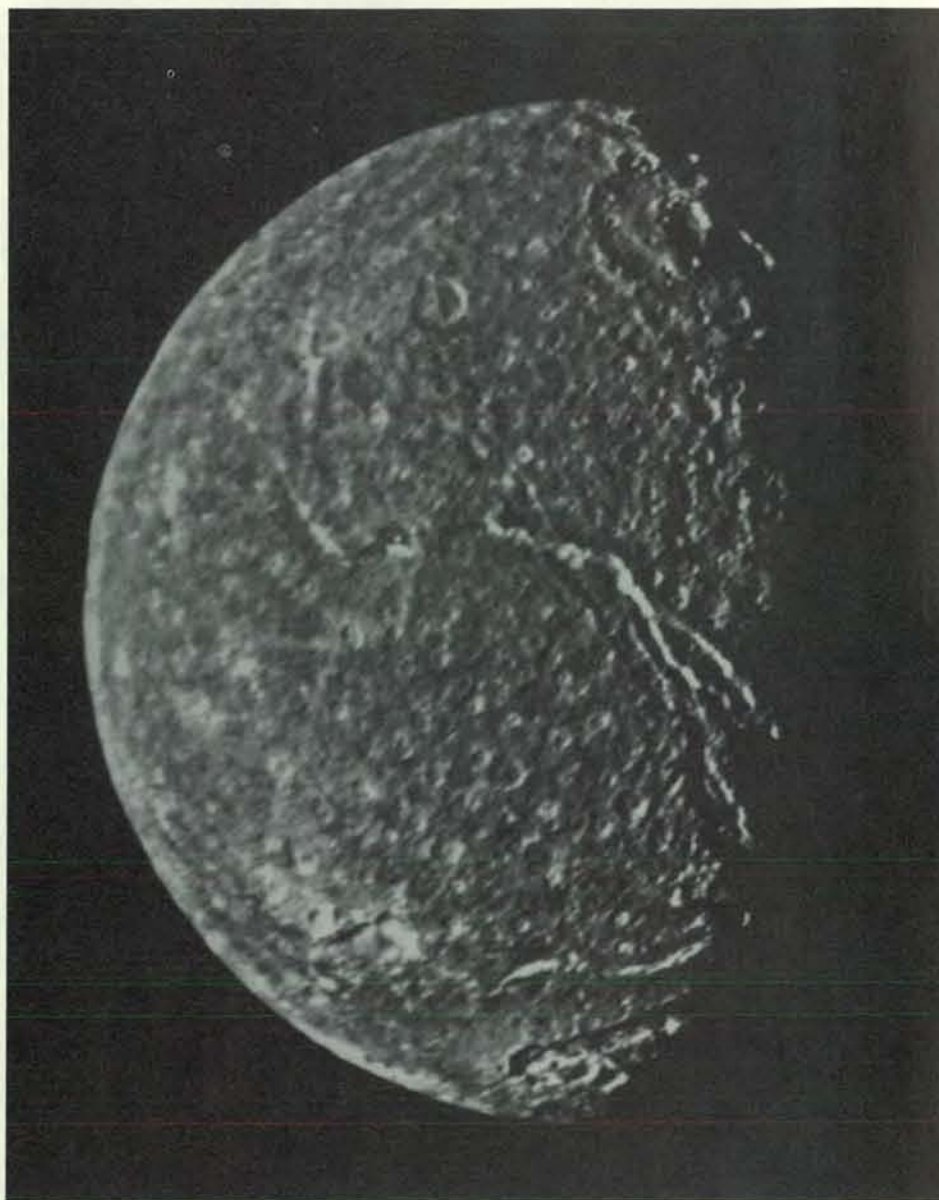
Tuttavia, ci sono problemi anche nell'ambito di questa ipotesi. Sebbene, infatti, la maggior parte di Umbriel sia uniforme, esso non è del tutto privo di particolari. Al contrario, possiede due configurazioni brillanti piuttosto evidenti nei pressi dell'equatore: un anello di 80 chilometri di diametro che sembra coprire il fondo di un cratere da impatto, e una macchia sul picco centrale di un altro grande cratere. L'origine di queste strutture è un mistero, ma il materiale brillante deve provenire dal sottosuolo. Questo indica che il rivestimento scuro, se esiste, deve essere sottile o assente in quelle due zone.

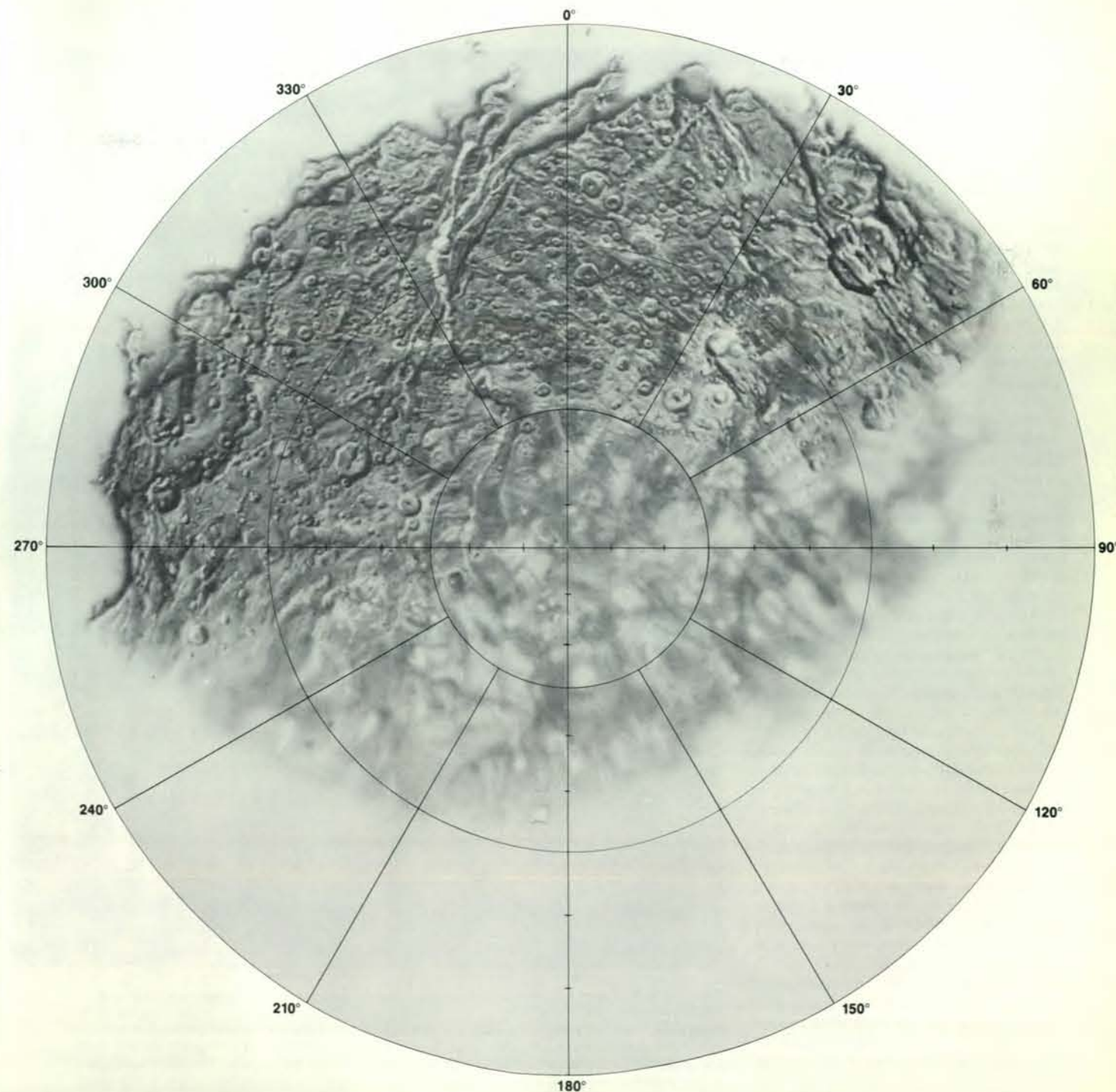
Mentre Umbriel è il più scuro dei satelliti principali di Urano, Ariel è il più brillante: ha un fattore di riflessione vicino al 40 per cento. E mentre la superficie di Umbriel è tra le più antiche e più craterizzate del sistema di Urano, quella di Ariel è una delle più giovani e meno craterizzate. La storia di Ariel ricorda quella di Titania, tranne per il fatto che l'attività geologica su Ariel è stata più intensa, più estesa e più prolungata nel tempo. Come Titania, Ariel non ha quasi crateri derivanti dal bombardamento di meteoriti della prima popolazione. Inoltre, ha solo circa un terzo del numero di crateri derivanti dal bombardamento di oggetti della seconda popolazione rispetto a Titania, il che indica che la sua superficie è stata rimodellata su un'estensione maggiore e per un periodo di tempo più lungo. Anche la struttura globale delle sue faglie di distensione ha uno sviluppo più completo; in certe zone di Ariel le fosse tettoniche sono profonde decine di chilometri.

Il rimodellamento della superficie di Ariel è stato un processo vulcanico. Le prove dell'estrusione di materiale sulla superficie sono molto più forti che nel caso di Titania. Il fondo delle fosse tettoniche nell'emisfero che è rivolto costantemente verso Urano (i satelliti di Urano, come la Luna, si trovano su orbite sincrone) è rivestito di uno strato uniforme e relativamente non craterizzato. Dato che questo strato ricopre le strutture geologiche preesistenti - l'esempio più evidente è un grande cratere semisepolto - esso deve essere derivato da colate vulcaniche.

In alcune aree questo strato uniforme mostra solchi e creste paralleli agli assi delle fosse tettoniche. Questo andamento suggerisce che il materiale sia fuoriuscito attraverso fratture disposte lungo gli assi e si sia sparso sul fondo degli avvallamenti, in maniera simile alle eruzioni di lava che sgorgano da fenditure presenti nel fondo degli oceani terrestri. Il materiale eruttato su Ariel quasi certamente non fu roccia fusa. Si trattò probabilmente di una miscela fluida, relativamente calda, di ghiaccio e roccia che ebbe un moto di scorrimento simile a quello di un ghiacciaio terrestre. Questa miscela deve essere stata piuttosto viscosa: in alcuni luoghi nei quali ha seppellito strutture superficiali più antiche, i suoi bordi formano una ripida scarpata di circa un chilometro di altezza.

Per quanto Ariel si sia rivelato un piccolo satellite sorprendentemente attivo, per certi versi il suo tipo di attività - fratturazione della crosta e colate vulcaniche ghiacciate - è alquanto familiare, simile a quello osservato nel caso dei satelliti di Giove, Europa e Ganimede, e di quello di Saturno, Encelado. Miranda, da parte sua, è invece uno dei mondi più strani che siano mai stati visti. Il più interno (e, con un diametro di circa 500 chilometri, il più piccolo) dei satelliti principali di Urano, ha a malapena abbastanza forza gravitazionale per assumere forma sferica. Eppure, la sua superficie è un guazzabuglio di formazioni geologiche complesse e peculiari che sembrerebbero più appropriate su un pianeta 10 volte più grande.

Le unità geologiche più antiche di Miranda sono pianure ondulate densamente craterizzate, «butterate» dal bombardamento di meteoriti della seconda popolazione. Esse coprono la maggior parte della superficie visibile nelle immagini di *Voyager 2* e non sembrano particolarmente strane. Sovrapposte alle pianure ondulate, però, ci sono tre enormi regioni di forma grossolanamente ovale che, in mancanza di un termine migliore, chiameremo regioni ovoidali. Esse hanno diametri tra 200 e 300 chilometri e, a giudicare dal numero di crateri che contengono, sono notevolmente più giovani delle pianure. Sono formate da fasce parallele di creste, solchi e scarpate che si appoggiano le une alle altre con strane angolazioni. Lungo le scarpate e nei crateri recenti all'interno delle regioni ovoi-

Titania è stato molto più attivo di Oberon dal punto di vista geologico. La fotografia è la composizione di due immagini riprese da *Voyager 2* da una distanza di 369 000 chilometri. Titania ha un minor numero di grandi crateri rispetto a Oberon, il che indica che la sua superficie è più giovane e possiede una rete di faglie di distensione che dimostra che la sua superficie è stata sottoposta a tensioni. In certe aree, per esempio presso l'equatore a una longitudine di 30 gradi (*in basso a destra nella fotografia*), la superficie è relativamente poco craterizzata. Queste aree livellate possono essersi formate per estrusione vulcanica di ghiaccio e roccia lungo le faglie.

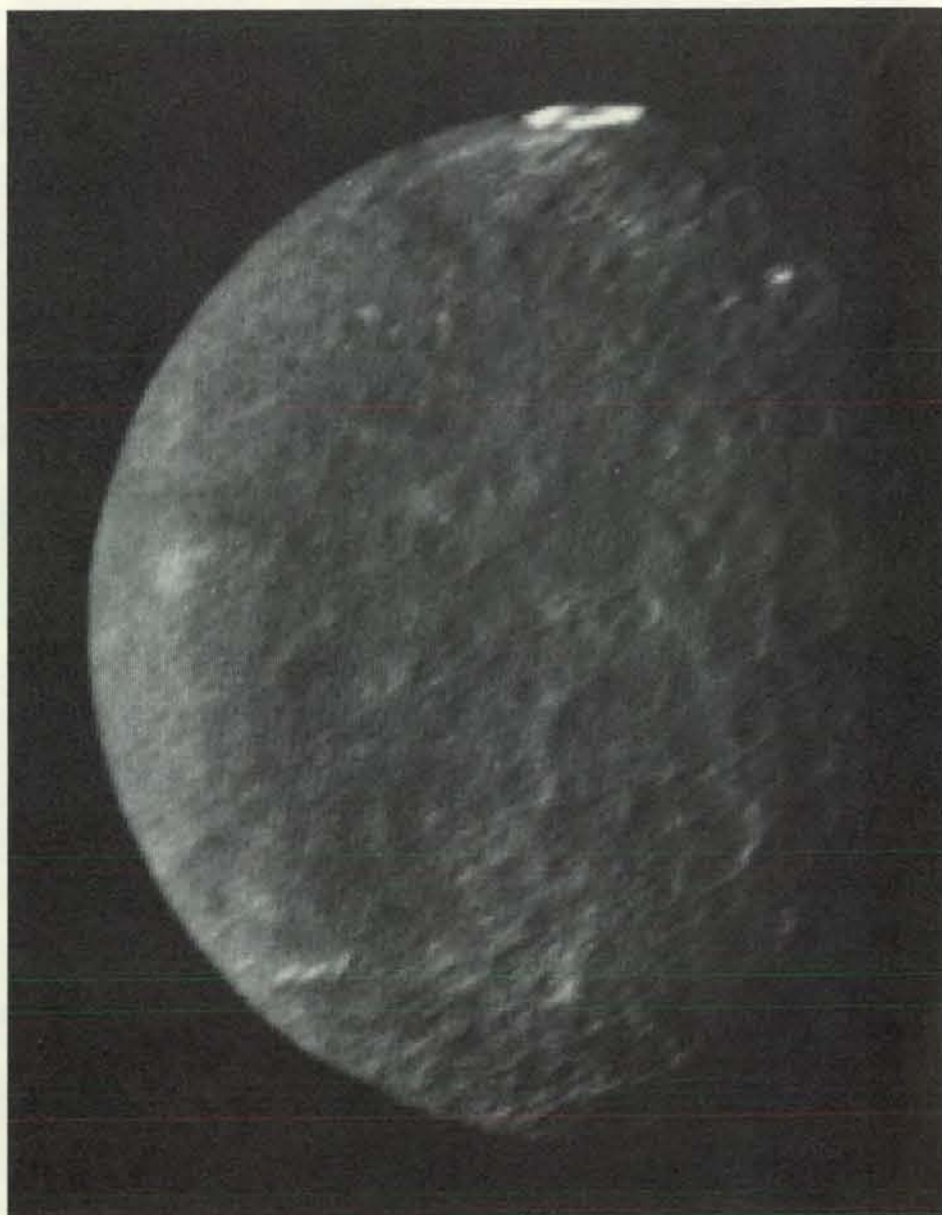
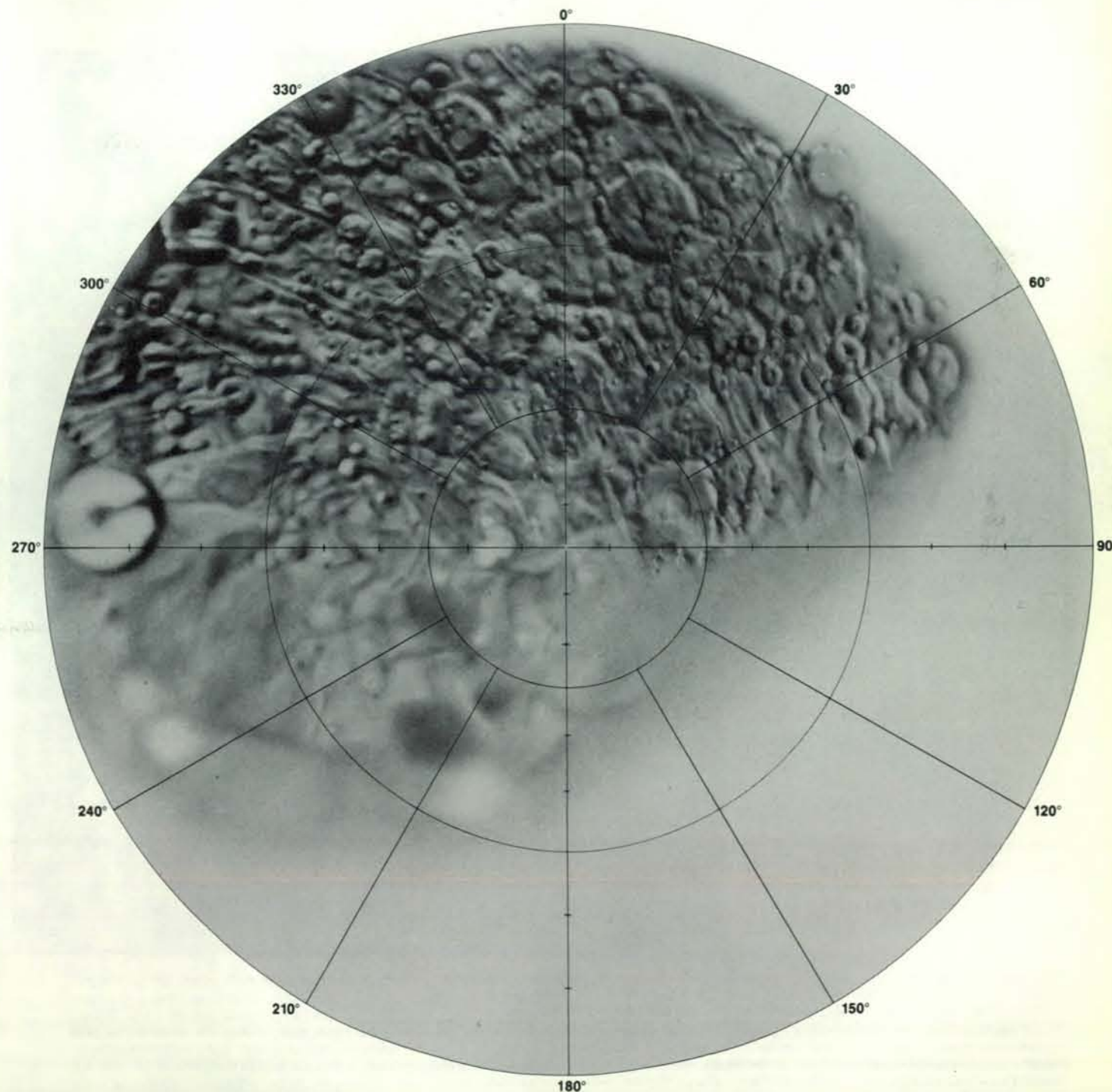
0°
330°
30°
300°
60°
270°
90°
240°
120°
210°
150°
180°

dali è visibile sia materiale brillante, sia materiale scuro; al contrario, le pianure ondulate hanno un fattore di riflessione piuttosto uniforme. Infine, sia le regioni ovoidali, sia le pianure sono solcate da enormi zone di frattura che compiono il periplo completo di Miranda, creando fosse tettoniche con ripidi fianchi terrazzati e con altezze che raggiungono anche 10 - 20 chilometri. E' probabile che il materiale delle regioni ovoidali sia meno denso di quello delle pianure. Le aree brillanti lungo le scarpate possono essere costituite da ghiaccio pulito proveniente dal sottosuolo, mentre il materiale scuro è probabilmente simile a quello eruttato dai crateri di Oberon. C'è qualche indizio di colate eruttive su Miranda, anche se esse sono molto meno estese rispetto a quelle di Ariel. In una delle regioni ovoidali, sembra che parte del materiale sia colato fuori da un basso cono a forma di ferro di cavallo, seppellendo scarpate e creste e accumulandosi dietro un ripido fronte lavico a forma di lobo.

I due modelli che sono stati proposti per spiegare le peculiarità di Miranda partono entrambi, per ragioni diverse, dal presupposto che le regioni ovoidali siano meno dense delle pianure ondulate. Il primo modello si basa sulle idee di Shoemaker riguardo all'andamento della craterizzazione primordiale. Shoemaker pensa che Miranda sia stato distrutto come Titania, riformandosi dopo ogni cataclisma, ma che ciò sia avvenuto cinque o più volte invece di una sola. Secondo la sua teoria, prima dell'ultimo evento distruttivo, la roccia e il ghiaccio del satellite ebbero il tempo di separarsi parzialmente. Quindi, quando Miranda fu nuovamente distrutto, si ruppe in grossi frammenti, ognuno dei quali era formato principalmente da ghiaccio o da roccia. Quando i frammenti si riaggregarono di nuovo in modo casuale, grossi pezzi di roccia, che erano vicino al nucleo del satellite distrutto, si vennero a trovare alla superficie del nuovo agglomerato. Il satellite così formatosi manteneva abbastanza calore da consentire al suo interno un moto viscoso. Di conseguenza, le masse di roccia presero a sprofondare ancora una volta verso il centro, consentendo al ghiaccio, più leggero, di salire e fluire nello spazio lasciato libero dalle masse rocciose,facendo sì che la superficie si fratturasse lungo linee di tensione concentriche. Le regioni ovoidali, in questo scenario, sono segni superficiali lasciati dallo sprofondamento delle masse di roccia.

A questa teoria si può contrapporre il fatto che difficilmente ammassi di roccia così grandi avrebbero potuto restare intatti durante la distruzione totale e la riaggregazione del satellite. Sembra più ragionevole pensare che simili cataclismi debbano aver dato origine a frammenti piccoli. Un punto di vista alternativo, proposto da uno di noi (Soderblom), ritiene che quando Miranda si aggregò l'ultima volta (a prescindere dal fatto che fosse stato davvero distrutto in precedenza oppure no), esso era costituito da una miscela abbastanza uniforme di roccia e ghiaccio. Quindi, sotto l'influenza del proprio calore interno, cominciò a differenziarsi. Mentre la roccia sprofondava verso il centro, agglomerati di ghiaccio presero a risalire in superficie. In questo scenario le regioni ovoidali non si formarono in seguito a grandi sprofondamenti di roccia, ma come conseguenza dell'ascesa di enormi masse di ghiaccio, che alla fine aprirono alcune brecce sulla superficie. La configurazione delle regioni ovoidali (che hanno tutte all'incirca le stesse dimensioni e sono regolarmente distanziate tra loro) suggerisce che il flusso fosse dotato di una qualche forma di organizzazione interna, simile forse a celle di convezione.

Sembra che il processo di differenziazione su Miranda non si sia mai completato. Se la differenziazione fosse continuata, le regioni ovoidali probabilmente sarebbero state cancellate e Miranda avrebbe finito per avere un mantello di ghiaccio solido intorno a un nucleo di roccia solida. Invece il satellite si arrestò a uno stadio embrionale di sviluppo geologico. Benché alcuni ricercatori ritengano che le regioni ovoidali possano avere un'età valutabile intorno a un solo miliardo di anni, a noi sembra più verosimile che l'evoluzione di Miranda si sia arrestata tre o quattro miliardi di anni fa.

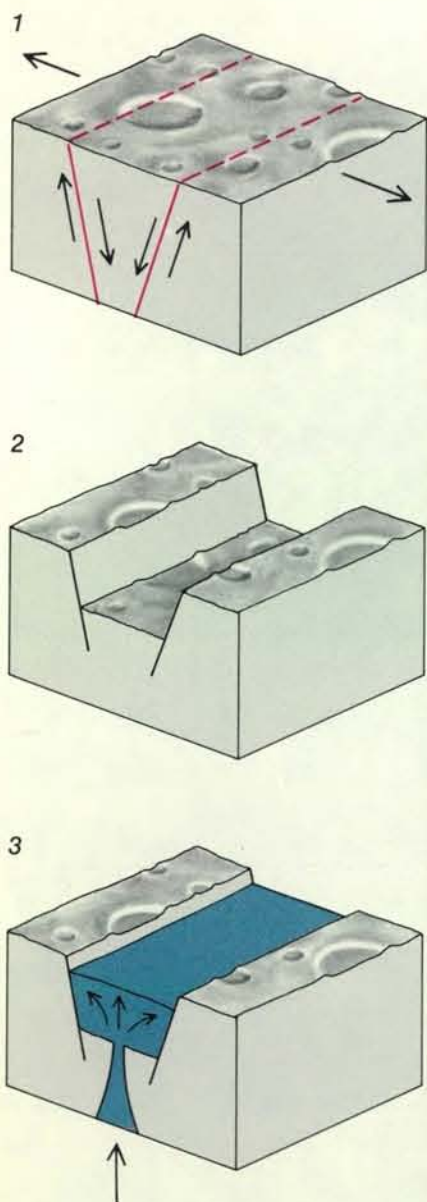
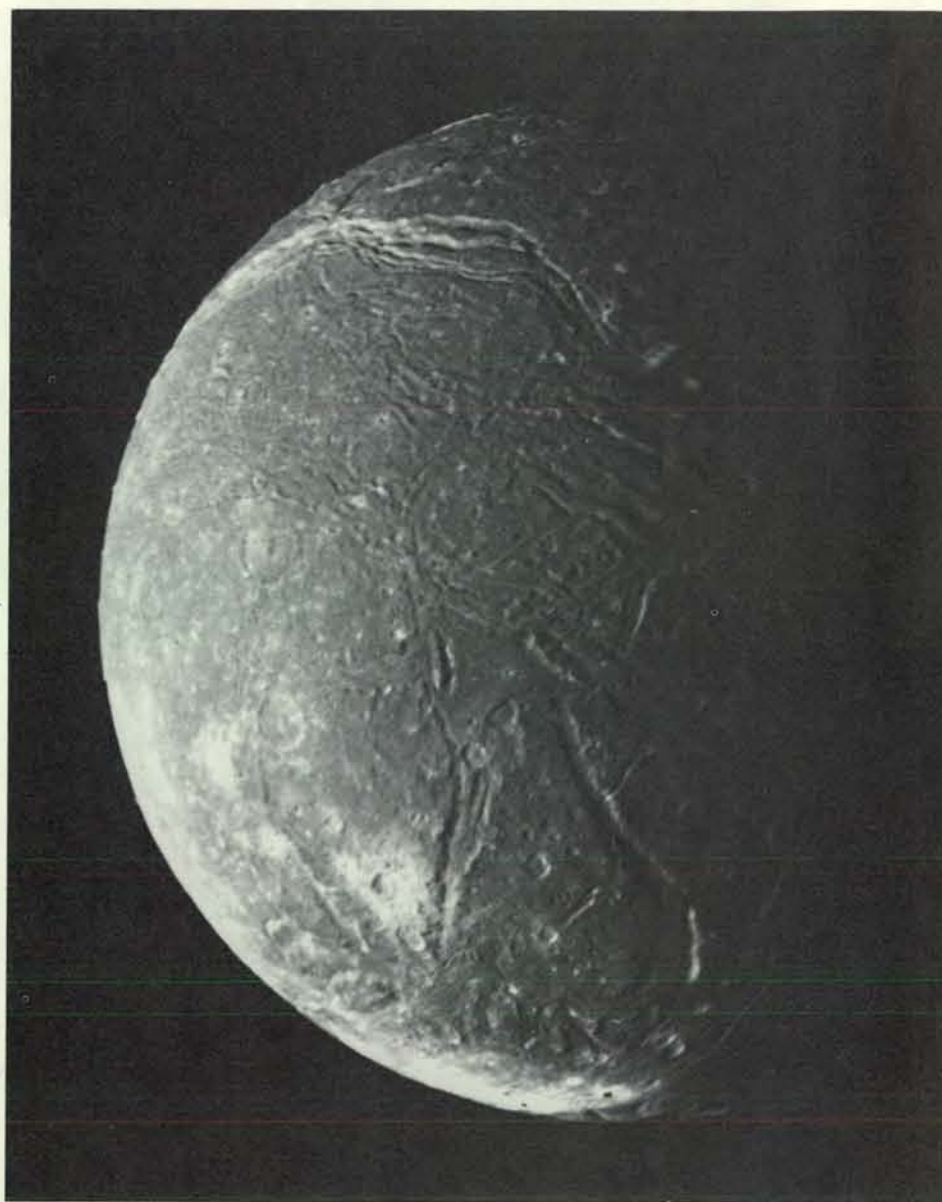
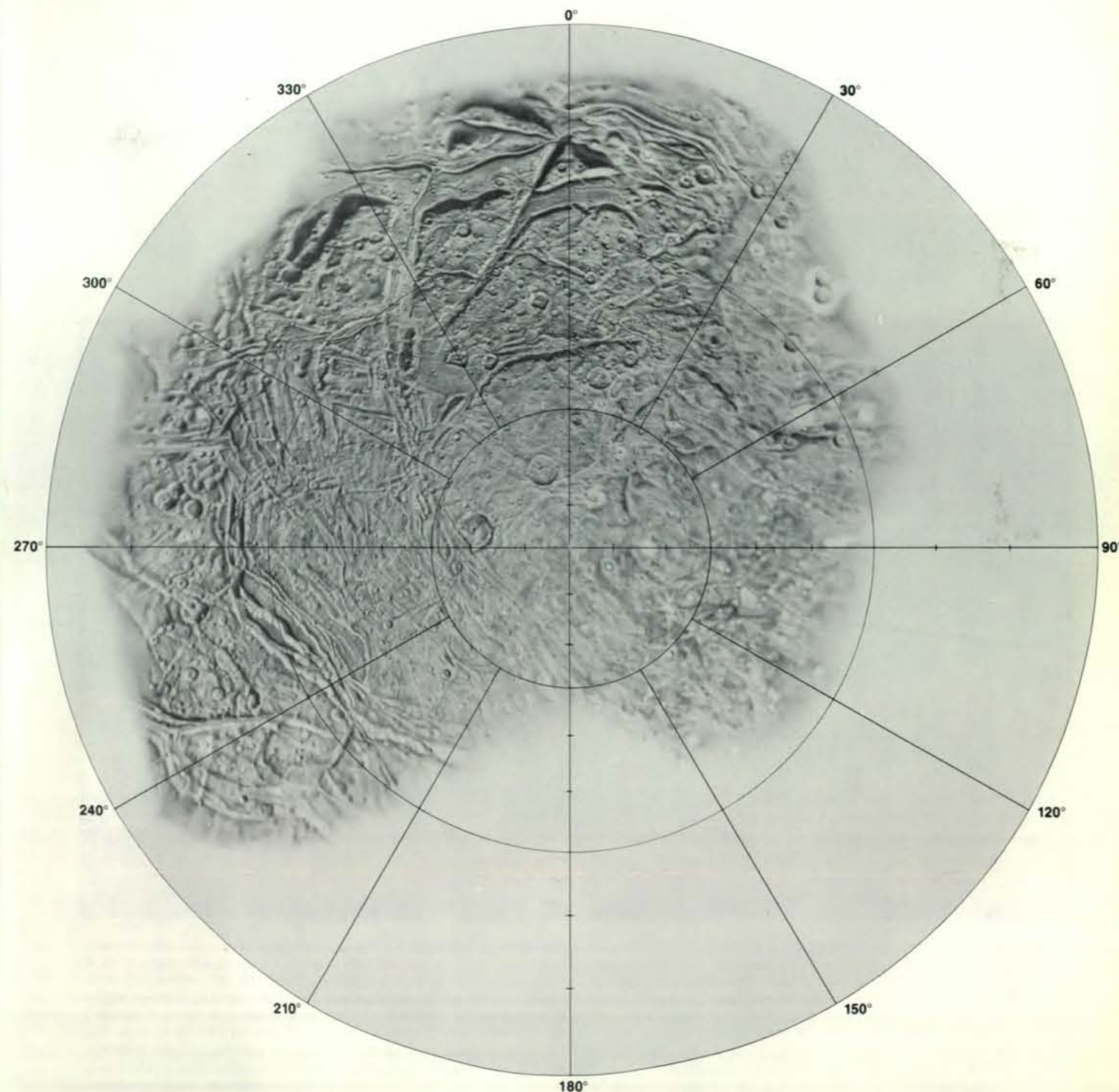
La cosa più sorprendente riguardo a satelliti come Ariel e Miranda non è il fatto che la loro attività geologica si sia arrestata molto tempo fa, ma soprattutto che essa sia potuta iniziare. Un tale grado di attività in corpi celesti così piccoli - Ariel è circa un terzo e Miranda meno di un sesto delle dimensioni della Luna - sarebbe sembrato straordinario dieci anni fa. Allora si pensava che la sorgente primaria di calore interno in un corpo planetario solido fosse il decadimento radioattivo di radionuclidi a lunga vita come uranio, potassio e torio. Dato che la quantità di elementi radioattivi in un corpo planetario è proporzionale al suo volume, mentre la velocità con cui perde calore è proporzionale alla sua superficie, un corpo grande conserva il calore interno molto più a lungo di uno piccolo, e quindi dà luogo ad uno sviluppo geologico più prolungato nel tempo. I pianeti di tipo terrestre si comportano secondo queste previsioni. La Terra è il più grande di questi pianeti ed è sempre stato chiaramente il più attivo dal punto di vista geologico; Marte è di dimensioni intermedie e ha avuto per tutta la sua storia un basso livello di attività vulcanica; Mercurio e la Luna sono piccoli e hanno avuto solamente una breve evoluzione primordiale.

I satelliti del sistema solare esterno non si adattano a questo scenario. Nel corso degli ultimi dieci anni le osservazioni da parte dei Voyager di più di 20 satelliti hanno portato a una drastica revisione delle ipotesi riguardanti la storia geologica dei corpi celesti di piccole di-

Umbriel si distingue soprattutto per la mancanza di caratteristiche: la sua superficie è di un grigio scuro pressoché uniforme e l'abbondanza di grandi crateri dimostra che è antica. Nella fotografia, ripresa da una distanza di 557 000 chilometri, si notano solo due particolari brillanti nei pressi dell'equatore. Il primo è un anello a una longitudine di 270 gradi (*in alto nella fotografia*) che sembra trovarsi in un cratere di 80 chilometri di diametro. Il secondo particolare è una macchia nel picco centrale di un altro cratere, a una longitudine di circa 310 gradi.

mensioni. Il satellite di Giove, Io, sul quale i vulcani sono tuttora in attività, è solo l'esempio più spettacolare; anche altri satelliti di Giove e Saturno mostrano i segni di un'attività geologica significativa. A questo elenco bisogna ora aggiungere Miranda, Ariel e forse anche Titania.

Le conseguenze delle osservazioni dei Voyager devono ancora essere ricavate in modo completo, ma hanno cominciato a emergere alcune idee-guida. I satelliti devono possedere un'altra sorgente di energia oltre ai radionuclidi a lunga vita. Può darsi che essi abbiano ricevuto un riscaldamento iniziale da parte di isotopi a vita breve, come lo iodio 129 o l'alluminio 26, che si pensa fossero presenti entrambi nella nebulosa solare primordiale. Un'altra possibile sorgente di riscaldamento è l'energia gravitazionale liberata dai frammenti che si andavano accumulando durante il processo di accrezione dei satelliti.

Le risonanze mareali sono una terza e potenzialmente più prolungata sorgente di riscaldamento. Quando i periodi orbitali di satelliti adiacenti differiscono di un numero intero, un satellite perturba l'orbita dell'altro a intervalli regolari che sono multipli del periodo del satellite perturbato. L'effetto è simile a quello che si ottiene sollecitando un pendolo secondo la sua frequenza di risonanza: la perturbazione si amplifica enormemente. Come conseguenza del fenomeno, l'orbita del satellite perturbato diventa nettamente ellittica. A causa dell'ellitticità, il rigonfiamento mareale provocato sulla superficie del satellite dalla forza di gravità del pianeta tende a spostarsi su e giù e avanti e indietro. Lo spostamento del rigonfiamento fornisce una quantità enorme di energia al satellite a causa dell'attrito e questa energia deve essere dissipata in calore. Si pensa che il riscaldamento mareale possa spiegare l'attività geologica su Io, Europa ed Encelado.

I satelliti di Urano non sono attualmente in risonanza. Ma, come è stato fatto notare da Steven W. Squyres, Ray T. Reynolds e Jack Lissauer della NASA in un lavoro pubblicato prima dell'attraversamento del sistema di Urano da parte di *Voyager 2*, Miranda, Ariel e Umbriel possono aver sperimentato una fase di risonanza nel loro passato geologicamente recente. Attualmente il riscaldamento mareale sembra essere la più probabile causa dell'intensa attività geologica di Miranda e Ariel. L'effetto mareale su Umbriel dovrebbe essere meno significativo dato che quest'ultimo si trova a una maggiore distanza da Urano; questo spiegherebbe perché Umbriel e Ariel si siano evoluti in maniera così diversa nonostante abbiano all'incirca le stesse dimensioni.

L'attività inaspettata dei satelliti nel sistema solare esterno deve avere qualcosa a che fare con la loro composizione chimica. Per dare origine a colate vulcaniche, infatti, un satellite deve non solo possedere una sorgente di calore interno, ma anche essere fatto di materiale che può fondere almeno parzialmente. I modelli della composizione chimica della nebulosa solare primordiale indicano che i satelliti del sistema solare esterno dovrebbero essere costituiti da rocce carbonacee e ghiacci, in primo luogo ghiaccio d'acqua. La temperatura superficiale dei satelliti di Urano è intorno agli 80 kelvin (−193 gradi Celsius); il ghiaccio d'acqua fonde a 273 kelvin. Qualunque sia stata la sorgente di calore interno dei satelliti, i calcoli teorici fanno ritenere molto improbabile che, in corpi così piccoli, la differenza tra la temperatura

Ariel mostra maggiori tracce di attività vulcanica avvenuta in tempi passati rispetto a tutti gli altri satelliti di Urano. La fotografia è un mosaico di immagini riprese da una distanza di circa 130 000 chilometri. Le caratteristiche più evidenti sono i *graben* (fosse tettoniche) che si incrociano presso l'equatore (*a destra nella fotografia*). Essi sono rivestiti da un materiale uniforme, poco craterizzato, che è fuoriuscito lungo il loro asse riempiendo il fondovalle. Il processo, comune sulla Terra, è mostrato a sinistra. Lo stiramento della crosta provoca la sua rottura lungo faglie di distensione (*1*); in seguito, un blocco di crosta si abbassa lungo le faglie, formando una fossa tettonica (*2*); infine, materiale parzialmente fuso risale dall'interno del satellite fino alla frattura (*3*). La quasi totale assenza di grandi crateri dimostra che la maggior parte della superficie di Ariel è stata ricoperta al termine del bombardamento da parte dei meteoriti che orbitavano intorno al Sole.

interna e quella esterna abbia mai superato i 200 gradi.

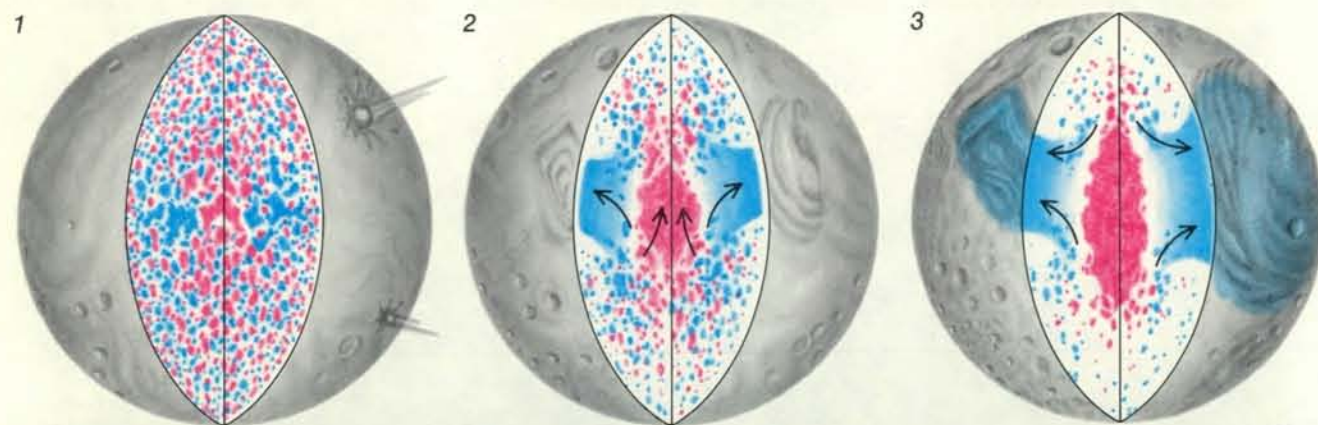
I satelliti probabilmente contengono qualche materiale che fonde a temperatura più bassa del ghiaccio d'acqua puro. I principali candidati sono il ghiaccio ammoniacale e i clatrati (composti di addizione formati per inclusione di molecole nelle cavità di un reticolo cristallino di una macromolecola) di ghiaccio d'acqua con metano o con ossido di carbonio. Una miscela di acqua e ammoniaca a saturazione, per esempio, ha un punto di fusione inferiore di circa 100 gradi rispetto a quello dell'acqua pura.

Le proporzioni relative di ghiaccio e roccia in un satellite influenzano anche la sua tendenza ad avere un'attività geologica. Più il satellite è roccioso, più è intensa la sua sorgente di calore di origine radioattiva, dato che gli elementi radioattivi sono contenuti nella roccia. Per valutare la quantità di roccia in un satellite bisogna conoscerne la densità media e per valutarne la densità bisogna sapere con una certa precisione qual è la sua massa. I calcoli della massa dei satelliti di Urano fatti prima della missione di *Voyager 2* si sono rivelati poco accurati, ma la sonda è stata in grado di misurare la massa di Oberon e Titania con precisione. (La procedura di misurazione è di per sé interessante. Nelle settimane precedenti l'incontro, i satelliti sono stati fotografati senza interruzione, dato che la conoscenza delle loro posizioni esatte era necessaria per dirigere la sonda. Le fotografie hanno evidenziato i minuscoli effetti di perturbazione reciproca tra i satelliti e queste perturbazioni, insieme alla esatta misurazione radio della posizione della sonda, hanno reso possibile il calcolo delle masse dei satelliti.)

Quando si passa dalla misura della massa a quella della densità, tenendo conto degli effetti di autocompressione gravitazionale del satellite, si constata che sia Oberon, sia Titania sono costituiti da materia la cui densità media varia tra 1,4 e 1,7 grammi per centimetro cubo. Questi valori implicano una percentuale di roccia compresa tra 40 e 65 per cento, il che li rende considerevolmente più rocciosi di Giapeto e di Rea, due satelliti di Saturno di dimensioni simili. Anche gli altri satelliti di Urano potrebbero contenere percentuali altrettanto alte di roccia. Questo suggerisce che il riscaldamento di origine radioattiva può aver dato un contributo significativo alla loro attività geologica.

La maggiore percentuale di roccia dei satelliti di Urano rispetto a quelli di Saturno contraddice quanto ci si dovrebbe attendere intuitivamente; in parole povere, sembrerebbe verosimile trovare più ghiaccio a mano a mano che ci si allontana dal Sole. Alcuni modelli della nebulosa solare avevano previsto che i satelliti di Urano non contenessero più del 35 per cento di roccia. A quanto pare, l'ipotesi era sbagliata.

Sono state proposte varie modifiche dei modelli. Una è basata sulla supposizione che le reazioni chimiche che hanno prevalso nelle fredde regioni esterne della nebulosa solare siano state diverse da quelle delle regioni interne. Un'altra, suggerita da James B. Pollack dell'Ames Research Center, parte dal presupposto che i satelliti di Urano abbiano acquisito gli elementi che costituiscono le rocce non dalla nebulosa solare, ma da un inviluppo gassoso legato a Urano. David J. Stevenson del California Institute of Technology ha avanzato l'ipotesi che un inviluppo ricco di materiali rocciosi potrebbe essere derivato dall'evento che si pensa abbia dato a Urano la sua orientazione fuori dal comune: l'impatto, durante gli stadi finali di accrezione del pianeta, con un corpo di dimensioni simili a quelle della Terra. Eventuali satelliti sarebbero probabilmente andati distrutti da un simile impatto, perciò quelli attuali dovrebbero essersi formati in seguito. In questo processo essi potrebbero avere incorporato materiali rocciosi provenienti dal corpo precipitato o dal pianeta stesso.

Le immagini di *Voyager 2* hanno dato molte informazioni sui satelliti di Urano, ma non sono riuscite a spiegare in modo chiaro perché essi siano scuri, rocciosi e geologicamente differenziati; questi problemi impegneranno i teorici. Noi stiamo già pensando a Nettuno, al suo satellite Tritone e al breve, ma importante incontro di *Voyager 2* nel 1989.

Miranda appare qui sotto un'angolazione dalla quale non è stata mai ripresa da *Voyager 2*. Dato che la sonda è passata molto vicino al satellite (entro 29 000 chilometri), la scala delle sue marcate configurazioni topografiche ha potuto essere dedotta attraverso immagini stereoscopiche della stessa regione riprese da angolazioni diverse. Miranda è il solo satellite del sistema solare esterno per il quale siano disponibili informazioni topografiche così dettagliate. Usando questi dati, i ricercatori dell'US Geological Survey hanno riproiettato le immagini di *Voyager 2* con nuove angolazioni per far risaltare le varie caratteristiche della superficie di Miranda. Le immagini rielaborate confermano la teoria secondo la quale questo satellite si arrestò a uno stadio embrionale di differenziazione. In base a questo modello, Miranda aveva sufficiente calore interno agli albori della sua storia perché si formassero agglomerati di roccia e ghiaccio (*1*). Successivamente, la roccia cominciò a sprofondare verso il nucleo e il ghiaccio prese a risalire in superficie (*2*). Infine, grossi «pennacchi» di ghiaccio aprirono alcune brecce in superficie formando le regioni ovoidali (*3*). Il calore interno di Miranda si esaurì prima che il processo di differenziazione fosse terminato.

Come i fotorecettori rispondono alla luce

Recenti ricerche sulla trasformazione dell'energia luminosa in segnali nervosi dimostrano che una singola cellula fotorecettrice della retina è in grado di registrare l'assorbimento di un solo fotone o quanto di luce

di Julie L. Schnapf e Denis A. Baylor

Il processo visivo inizia con la conversione di pacchetti discreti di energia elettromagnetica, chiamati fotoni o quanti di luce, in segnali nervosi che il cervello è in grado di analizzare. Questo processo viene eseguito dalle cellule fotorecettrici dell'occhio che sono disposte a mosaico sulla superficie posteriore della retina, mentre i neuroni rivestono la parte interna del bulbo oculare. La cornea e il cristallino dell'occhio inviano un'immagine del mondo esterno sullo strato di fotorecettori. Ciascuna cellula assorbe la luce da un unico punto dell'immagine e genera un segnale elettrico che codifica la quantità di luce assorbita. Questi segnali vengono trasmessi attraverso un complicato sistema di sinapsi, o giunzioni nervose, che si trovano nella retina e nel cervello. A livello di queste giunzioni i segnali provenienti dai fotorecettori sono sommati e confrontati. Questo processo permette al sistema visivo di ottenere informazioni sulla forma, il movimento e il colore del mondo circostante.

Dato il ruolo fondamentale dei recettori nella visione, è sorprendente che il loro meccanismo di azione sia rimasto sconosciuto per lungo tempo, ma la situazione si è profondamente modificata nel corso dell'ultimo quarto di secolo. Infatti i metodi sempre più perfezionati messi a punto per registrare l'attività elettrica dei singoli fotorecettori hanno fornito informazioni dettagliate sul meccanismo di conversione dell'energia luminosa in segnali nervosi. Le nuove tecniche hanno consentito di osservare direttamente il segnale connesso con l'assorbimento di un solo fotone e, sulla base di queste misurazioni, si sono potute avanzare semplici spiegazioni riguardanti alcuni aspetti dell'intera percezione visiva: si è compreso perché captiamo gli stimoli deboli più lentamente di quelli intensi, perché vediamo talvolta luce in condizioni di buio completo e perché, in alcuni casi, una miscela di radiazioni di svariate lunghezze d'onda ci dà la stessa sensazione cromatica di una radiazione di una sola lunghezza d'onda.

L'occhio della maggior parte dei vertebrati contiene due tipi di fotorecettori: i coni e i bastoncelli. Questi ultimi consentono la visione in luce debole, ma sono tanto sensibili da soprasaturarsi alla normale luce diurna, diventando incapaci di trasmettere segnali. I coni consentono la visione alla luce del giorno, in quanto operano efficientemente a elevate intensità luminose, fornendo una visione più ricca di dettagli temporali e spaziali e permettendo la percezione dei colori.

Coni e bastoncelli contengono organelli specializzati per la trasduzione e la trasmissione dei segnali. A un'estremità della cellula (la più lontana dal cristallino) si trova il cosiddetto segmento esterno, che assorbe la radiazione luminosa e genera segnali elettrici. All'altra estremità si trova la terminazione sinaptica, che trasmette i segnali ad altre cellule nervose (cellule bipolari e orizzontali) della retina rilasciando un mediatore chimico. Tra il segmento esterno e la terminazione sinaptica si trova una regione detta segmento interno.

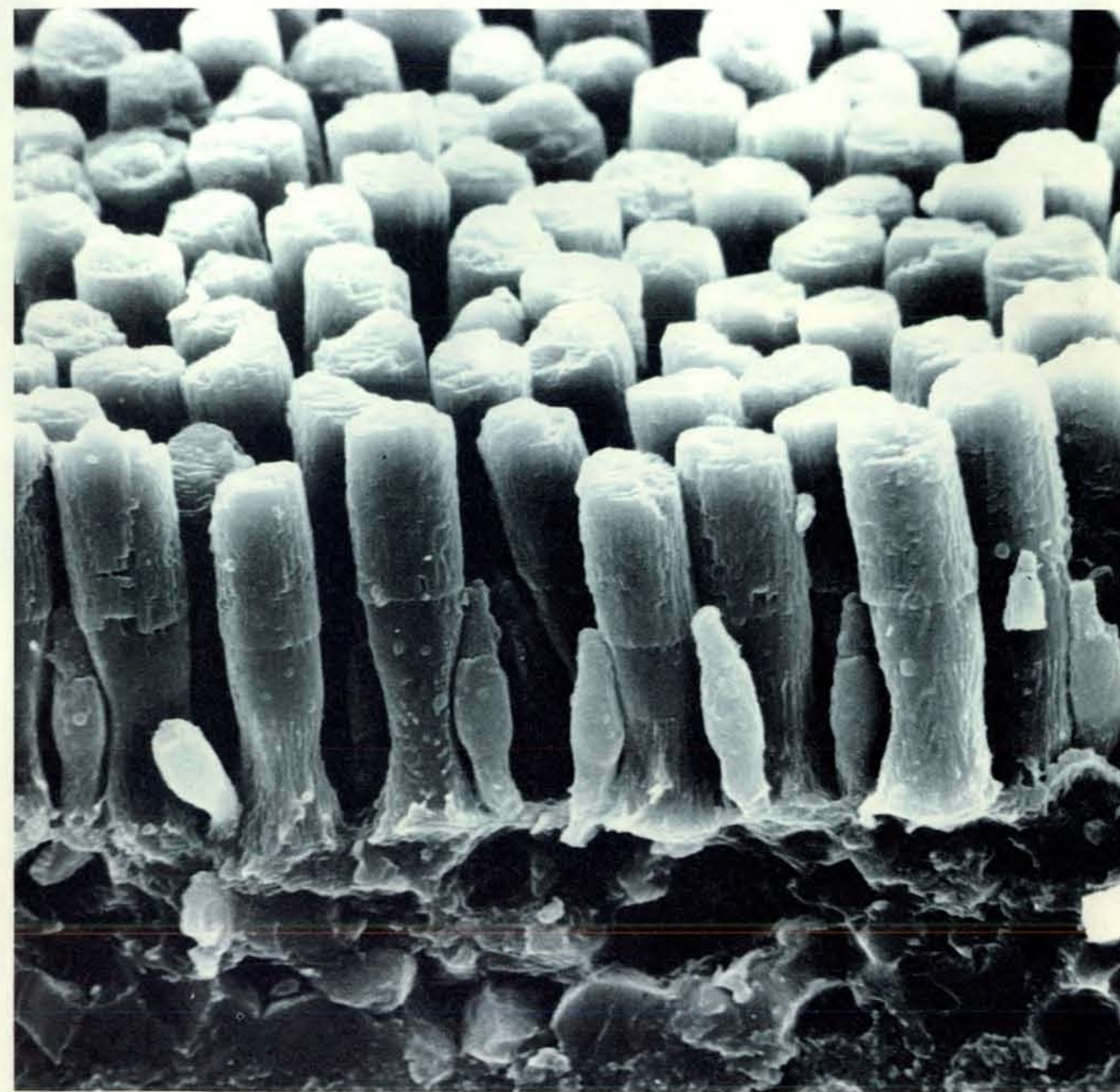
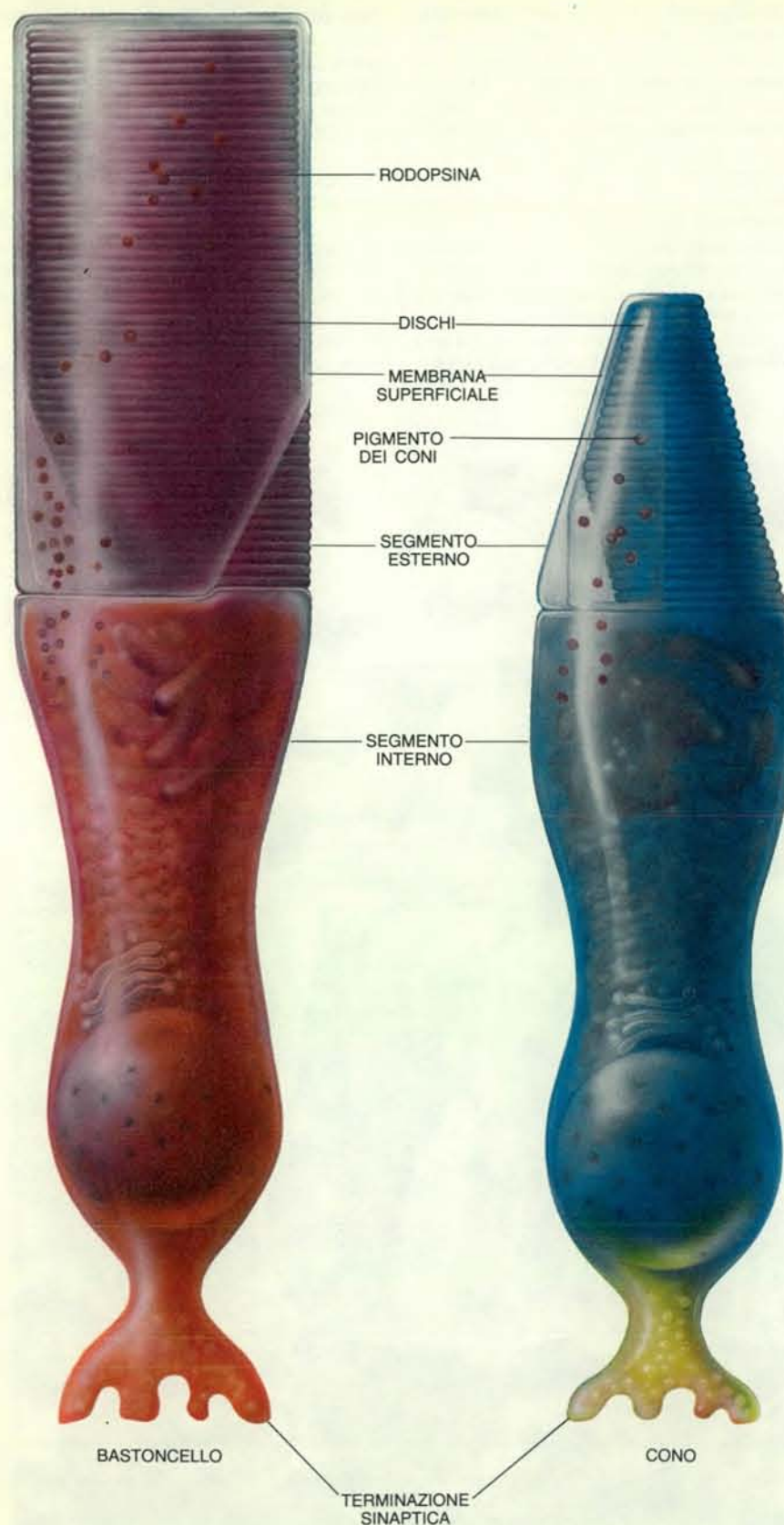
Il segmento esterno di un bastoncello è cilindrico, mentre quello di un cono è appuntito (da ciò derivano i nomi delle due cellule). Entrambi i tipi di segmento esterno sono formati da estese superfici sovrapposte di membrane rese fotosensibili dalle molecole di pigmento in esse contenute. I bastoncelli contengono il pigmento rosso rodopsina. Nella retina umana si trovano tre tipi di coni, ciascuno dei quali contiene un pigmento che mostra, nel campo di frequenze dello spettro visibile, il massimo assorbimento rispettivamente nella regione delle lunghezze d'onda corte, medie o lunghe. Queste differenze nelle bande di assorbimento dei tre pigmenti presenti nei coni costituiscono la base della visione dei colori: infatti, con illuminazione scarsa, quando il processo visivo è mediato dai bastoncelli, tutti gli oggetti appaiono privi di colore.

All'interno dei bastoncelli la membrana fotosensibile è disposta in una pila ordinata di dischi racchiusi entro una membrana superficiale separata, come una pila di monete dentro un contenitore cilindrico. Nei coni, invece, la membrana fotosensibile consiste di un solo grande strato ripiegato, che assume anche la funzione di membrana superficiale. Lo studio delle caratteristiche della membrana dei bastoncelli mostra che una sostanza diffusibile, «un mediatore interno», porta l'informazione dai dischi, dove viene assorbita la luce, alla membrana superficiale, dove viene generato il segnale elettrico. Ricerche condotte di recente in molti laboratori hanno consentito di identificare questo mediatore chimico in un nucleotide, il guanosinmonofosfato ciclico (cGMP), che partecipa anche alla trasduzione operata nei coni.

In quale modo l'assorbimento della luce da parte di un cono o di un bastoncello porta alla produzione di un segnale elettrico? Per rispondere a questa domanda occorre per prima cosa comprendere il comportamento dei fotorecettori al buio. Si potrebbe pensare che le cellule, in assenza di luce, siano inerti, ma in realtà esse sono più che mai attive. La membrana di un fotorecettore, come quella di ogni cellula, separa soluzioni con concentrazioni differenti di ioni (atomi dotati di carica elettrica). Le soluzioni che si trovano all'interno e all'esterno di un fotorecettore contengono ioni sodio e ioni potassio carichi positivamente. All'esterno della cellula la concentrazione di ioni sodio è elevata e quella di ioni potassio bassa, mentre, al contrario, all'interno della cellula la concentrazione degli ioni potassio è elevata e quella degli ioni sodio bassa. Queste differenze di concentrazione sono mantenute dall'attività di una pompa che utilizza l'energia metabolica per l'espulsione di ioni sodio e l'introduzione di ioni potassio.

In condizioni di riposo, le membrane della maggior parte dei neuroni mostrano maggior permeabilità agli ioni potassio, i quali quindi le possono attraversare liberamente, rispetto agli altri ioni. Siccome gli ioni potassio hanno concentrazione più elevata all'interno della cellula, tendono a diffondere attraverso la membrana verso l'esterno. A mano a mano che la diffusione procede, la carica si sposta dalla superficie interna della membrana a quella esterna e, a causa di questo trasferimento di carica, il potenziale interno diventa negativo rispetto a quello esterno, con una differenza caratteristica di 100 millivolt. In un fotorecettore la permeabilità agli ioni potassio è massima a livello del segmento interno e della terminazione sinaptica.

Al buio, un fotorecettore mostra una discreta permeabilità anche agli ioni sodio; questi, quindi, fluiscono dalla soluzione esterna più concentrata verso l'interno del segmento esterno, producendo una corrente elettrica diretta verso l'interno. Questa corrente è bilanciata dal movimento di ioni potassio verso l'esterno che si verifica nel resto della cellula: questo flusso di corrente è chiamato «corrente al buio».

Quando un cono o un bastoncello assorbe luce, il flusso del sodio verso l'interno si blocca, riducendo la «corrente al buio» e determinando l'aumento della polarizzazione negativa all'interno della cellula. Questa inversione

I coni e i bastoncelli della retina di una salamandra appaiono ingranditi 2000 volte in questa microfotografia realizzata al microscopio elettronico a scansione da Scott Mittman e Maria T. Maglio dell'Università della California a San Francisco. Le cellule cilindriche sono i bastoncelli, mentre quelle appuntite, più piccole, sono i coni. Le cellule fotorecettrici della retina dell'uomo sono circa quattro volte più sottili.

I coni e i bastoncelli, pur differendo per forma e funzione, mostrano alcune somiglianze. La parte superiore delle due cellule, o segmento esterno, contiene molecole di pigmenti che assorbono la luce; la parte inferiore, o segmento interno, contiene il nucleo e i mitocondri; la terminazione sinaptica collega i fotorecettori ad altre cellule della retina. I bastoncelli mediano la visione con luce debole, mentre i coni sono responsabili della visione alla luce del giorno. Nella retina umana esistono tre tipi di coni, ciascuno dei quali contiene un pigmento che assorbe fortemente nella regione del blu, del verde o del rosso dello spettro visibile, permettendo la visione dei colori.

del potenziale di membrana viene chiamata iperpolarizzazione, mentre con il termine «fotocorrente» si fa riferimento al fenomeno della riduzione della corrente al buio. I due tipi di corrente vennero descritti per la prima volta intorno al 1970 da William A. Hagins, Richard D. Penn e Shuko Yoshikami, che allora lavoravano presso il National Institute of Arthritis and Metabolic Diseases. Questi ricercatori effettuarono misurazioni del flusso di corrente in un gran numero di bastoncelli.

L'iperpolarizzazione determinata dalla luce è prodotta a livello del segmento esterno, ma si diffonde alla terminazione sinaptica, da dove viene trasmessa ad altre cellule della retina. L'iperpolarizzazione può essere registrata con un microelettrodo inserito in uno dei coni o dei bastoncelli relativamente grandi che si trovano nella retina di alcuni pesci, anfibi e rettili. I primi esperimenti di questo tipo vennero condotti alla metà degli anni sessanta in Giappone da Tsuneo Tomita e collaboratori presso la Keio University. Queste registrazioni dimostrano che, al buio, il potenziale di membrana è di circa -40 millivolt (mV). Un lampo di luce (flash) determina un'iperpolarizzazione che aumenta con l'intensità luminosa e, in particolare, quando esso è molto forte, la risposta raggiunge il valore limite di circa -30 mV, in corrispondenza del quale il potenziale di membrana è di -70 mV.

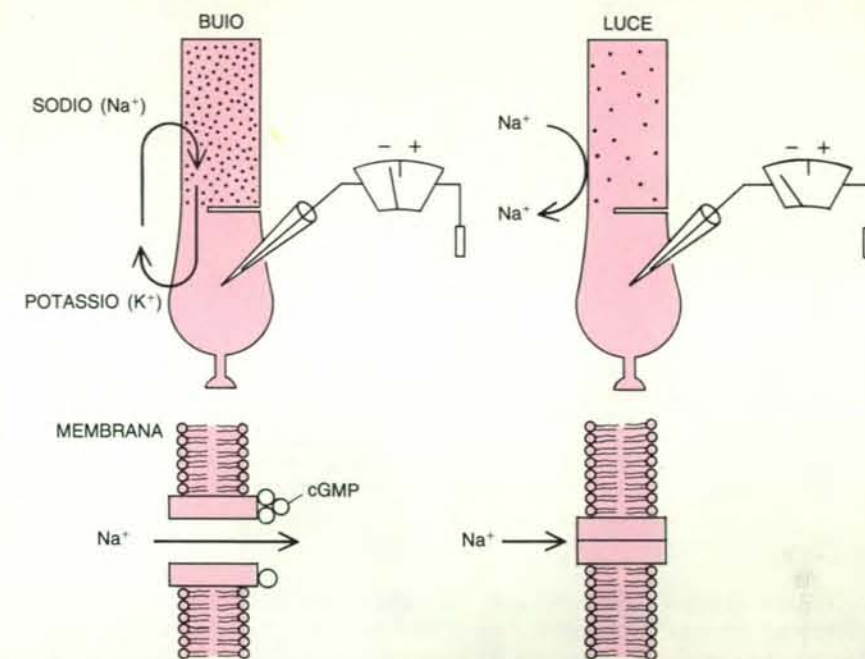
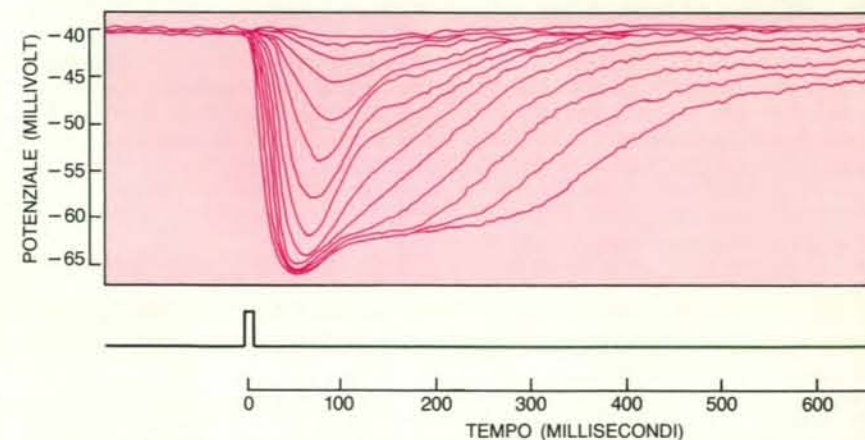
Nel 1975 Robert Fettiplace e uno degli autori (Baylor), che lavoravano presso la Stanford University School of Medicine, dimostrarono che l'iperpolarizzazione è in realtà necessaria e sufficiente per controllare il flusso di informazioni che arrivano ad altri neuroni visivi attraverso le sinapsi. Per simulare o impedire l'iperpolarizzazione prodotta dalla luce ci siamo serviti di un elettrodo intracellulare che faceva pervenire la corrente elettrica in un singolo fotorecettore. Contemporaneamente, abbiamo registrato le risposte di un'altra cellula, detta cellula gangliare, più lontana nella sequenza dei neuroni della retina. Siamo riusciti a riprodurre la risposta gangliare a un breve lampo luminoso proiettato nel fotorecettore, iperpolarizzando artificialmente il fotorecettore al buio. Inoltre, la cellula gangliare non rispondeva quando bloccavamo l'iperpolarizzazione provocata dalla luce inviando una corrente depolarizzante.

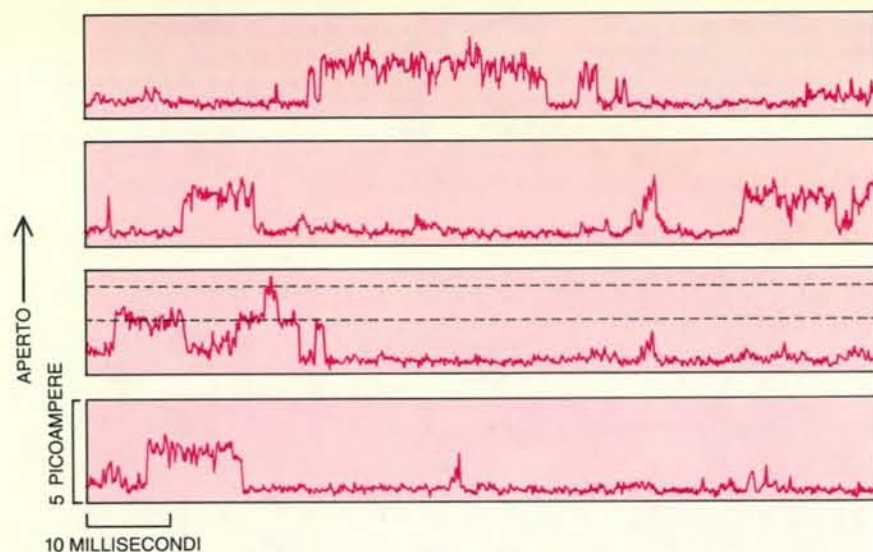
In quale modo l'assorbimento della luce blocca l'afflusso di ioni sodio a livello del segmento esterno? Al buio sia i coni, sia i bastoncelli contengono un'elevata concentrazione di guanosinmonofosfato ciclico, che è disposto intorno ai pori della membrana superficiale e li tiene aperti, consentendo agli ioni sodio di entrare. In condizioni di luce la concentrazione di cGMP diminuisce, il nucleotide lascia i siti di legame a livello dei pori che quindi si chiudono. A causa di questo fenomeno, la permeabilità della membrana agli ioni sodio diminuisce e la membrana stessa si iperpolarizza.

La serie di eventi molecolari che porta alla riduzione del cGMP si articola in tre passaggi. George Wald e colleghi della Harvard University dimostrarono alcuni anni fa che i pigmenti dei coni e dei bastoncelli contengono un componente capace di assorbire la luce chiamato 11-*cis* retinene, accoppiato a una proteina che «sintonizza» l'assorbimento con una particolare regione dello spettro visibile: le proteine del pigmento dei bastoncelli, la rodopsina, e dei tre pigmenti dei coni sono differenti. Quando il retinene presente nella rodopsina assorbe un fotone di luce cambia configurazione e fa sì che la parte proteica della molecola diventi enzimaticamente attiva. Come dimostrarono successivamente Lubert Stryer e collaboratori della Stanford University, la forma attivata della rodopsina attiva, a sua volta, molte molecole di una proteina che i ricercatori chiamarono trasducina. A loro volta, le molecole attivate di trasducina attivano un enzima che scinde il cGMP. Il sistema si comporta come un fotomoltiplicatore chimico: l'assorbimento di un solo fotone causa la rapida rottura di centinaia di molecole di cGMP e blocca l'ingresso di un milione di ioni sodio.

Recentemente, Evgeniy Fesenko e colleghi dell'Accademia delle Scienze dell'Unione Sovietica hanno ottenuto importanti risultati a sostegno dell'idea che il cGMP controlli in realtà il trasporto degli ioni sodio attraverso la membrana superficiale. Questi ricercatori applicarono una micropipetta, cioè un capillare di vetro del diametro di un micrometro (10^{-6} metri), alla membrana superficiale del segmento esterno di un bastoncello ottenuto dalla retina di rana. Aspirando delicatamente con la pipetta e allontanandola rapidamente strapparono il pezzetto di membrana che aderiva alla punta e scoprirono che, immergendolo in una soluzione contenente cGMP, questo diventava permeabile al sodio.

Il meccanismo molecolare alla base del movimento del sodio regolato dal cGMP fu comunque chiarito solo quando Anita Zimmerman del nostro laboratorio e Lawrence Haynes e King-Wai Yau della University of Texas Medical Branch di Galveston riuscirono a dimostrare che la corrente che attraversa un singolo sito di permeazione può superare il milione di ioni sodio al secondo. Questo flusso procede con una velocità di due ordini di grandezza superiore a quella con cui operano i trasportatori di membrana, che debbono subire un cambiamento di configurazione per trasferire uno o pochi ioni alla volta. Un flusso così rilevante significa invece che gli ioni attraversano la membrana diffondendo attraverso pori pieni d'acqua. Sembra che l'apertura di un singolo poro sia controllata contemporaneamente da tre o più molecole di cGMP; in altre parole, ciascun poro si comporta come un efficiente interruttore molecolare destinato a riconoscere cambiamenti infinitesimali della concentrazione del cGMP.

In opportune condizioni, un bastoncello della retina umana assorbe un singolo fotone che attiva una sola molecola di rodopsina, benché nel bastoncello vi siano circa cento milioni di molecole.

La risposta elettrica alla luce di un cono o di un bastoncello è dovuta a una riduzione della permeabilità agli ioni sodio nella membrana superficiale del segmento esterno. Al buio gli ioni sodio, carichi positivamente, fluiscono entro la cellula riducendo la densità di carica negativa all'interno della membrana celllulare (*in alto a sinistra*). Un flusso verso l'esterno di ioni potassio attraverso il segmento interno e la terminazione sinaptica completa un ciclo continuo di «corrente al buio». Responsabile del mantenimento dell'elevata permeabilità al sodio è il nucleotide guanosinmonofosfato ciclico (cGMP), la cui concentrazione è elevata al buio (*punteggiatura*); in condizioni di buio diverse molecole di cGMP disposte attorno a un poro ne determinano l'apertura (*in basso a sinistra*). Alla luce la concentrazione di cGMP diminuisce, il nucleotide lascia i siti di controllo e i pori si chiudono (*in basso a destra*). L'ingresso di ioni sodio quindi si blocca e il potenziale interno della cellula si iperpolarizza, cioè diventa più negativo (*in alto a destra*).

In questo esperimento sono state registrati i potenziali di iperpolarizzazione misurati inserendo un microelettrodo intracellulare in un cono sensibile al rosso appartenente a una retina di tartaruga. I tracciati sovrapposti nel grafico corrispondono alle risposte a brevi flash di intensità luminosa crescente. La differenza di potenziale attraverso la membrana è qui riportata in funzione del tempo trascorso dall'emissione del flash. Le intensità dei flash differivano di un fattore due: il più debole riusciva ad attivare circa 50 molecole del pigmento che assorbe la luce nei coni; i flash più intensi determinavano ampiezze di risposta corrispondenti alla saturazione. In quest'ultimo caso, il potenziale di membrana raggiungeva il valore di circa -65 millivolt.

APERTO
5 PICOAMPERE
10 MILLISECONDI

Un singolo poro del segmento esterno di un bastoncello dà origine, aprendosi, a una corrente elettrica misurabile. Nell'esperimento un pezzetto della membrana di un bastoncello di salamandra è stato immerso in una soluzione contenente cGMP, mentre il potenziale di membrana era mantenuto a +75 mV. Le deviazioni del tracciato verso l'alto corrispondono all'apertura di un singolo poro. Nel terzo grafico dall'alto si osserva l'apertura contemporanea di due pori. Per rendere più precisa la misura, la concentrazione degli ioni calcio e magnesio era molto bassa.

Questa capacità venne per la prima volta dimostrata nel corso di esperimenti condotti agli inizi degli anni quaranta da Selig Hecht, Simon Shlaer e Maurice H. Pirenne della Columbia University. Questi ricercatori diressero lampi di luce di debole intensità nell'occhio di un soggetto seduto in condizioni di buio completo e scoprirono che, se si variava l'intensità del lampo luminoso, il soggetto lo percepiva solo quando venivano assorbiti sette fotoni. Poiché una popolazione di 500 bastoncelli assorbiva i fotoni secondo uno schema spaziale casuale, era improbabile che ogni bastoncello avesse assorbito più di un fotone. Da questo fatto gli studiosi conclusero che un bastoncello deve produrre un segnale rilevabile anche quando assorbe un solo fotone.

Quali sono l'ampiezza e la forma del segnale elettrico generato dall'assorbimento di un singolo fotone? Con quale probabilità questo segnale si manifesta dopo un assorbimento? Segnali di questo tipo possono essere generati al buio? Nei primi anni settanta ricercatori di diversi laboratori cercarono di registrare la risposta dei bastoncelli in termini di differenza di potenziale dovuta all'assorbimento di un quanto di luce. Dapprima non ottennero alcun risultato, perché i bastoncelli «associano» i loro segnali: infatti connessioni specializzate, chiamate giunzioni facilitanti intervallate, collegano i bastoncelli che si trovano vicini nella retina, consentendo alle correnti elettriche di passare liberamente al loro interno. Di conseguenza, l'iperpolarizzazione di risposta a un singolo fotone viene ripartita tra 10 o più bastoncelli, divenendo troppo piccola per essere rilevata.

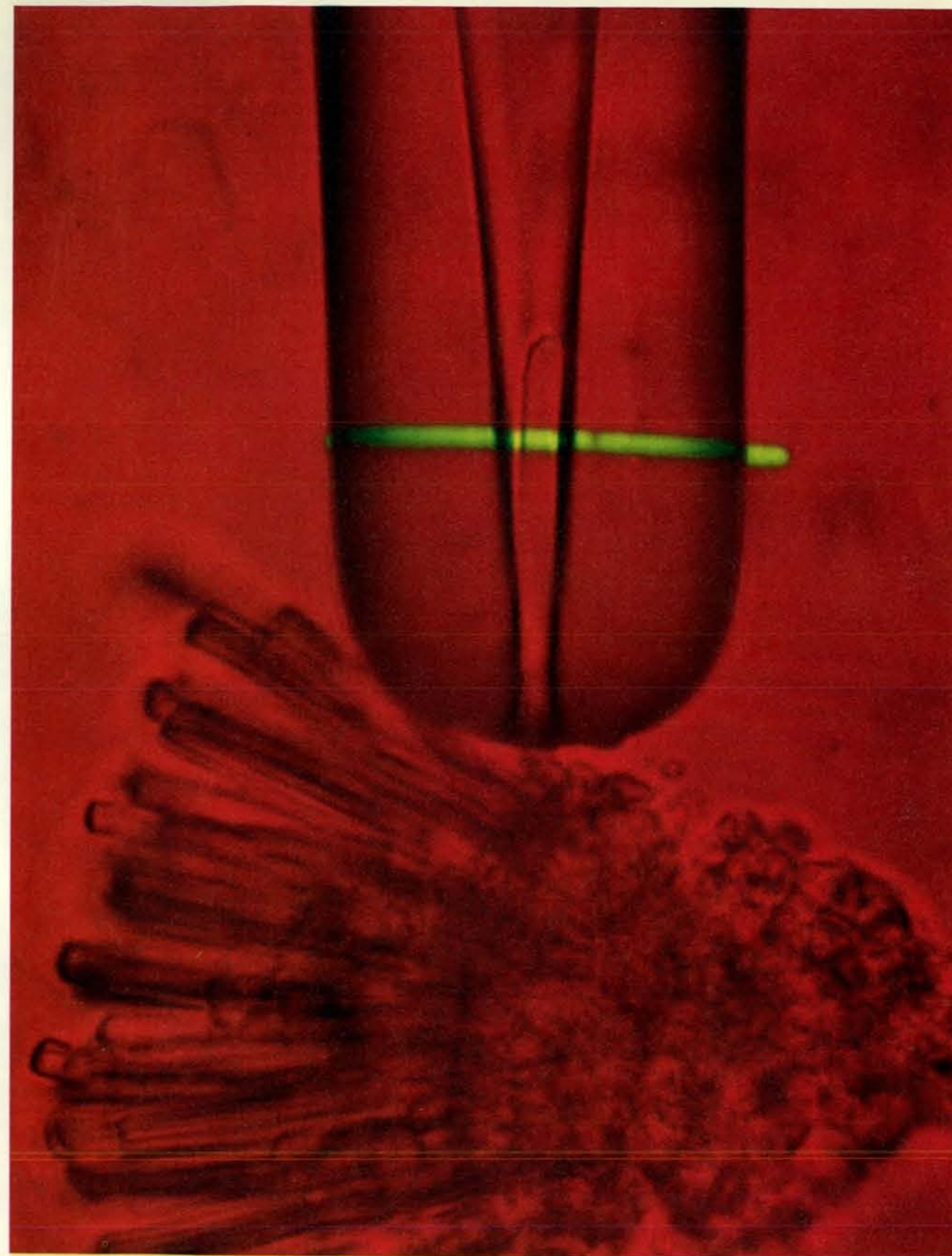
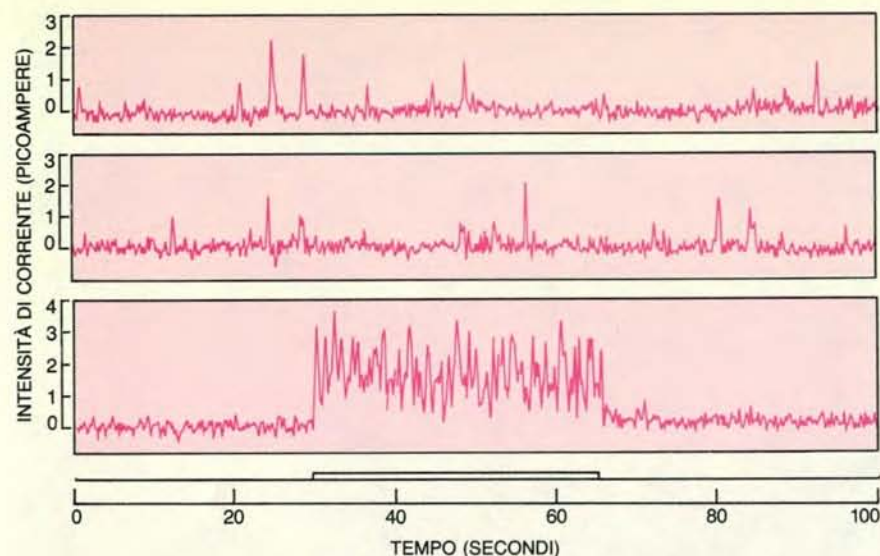
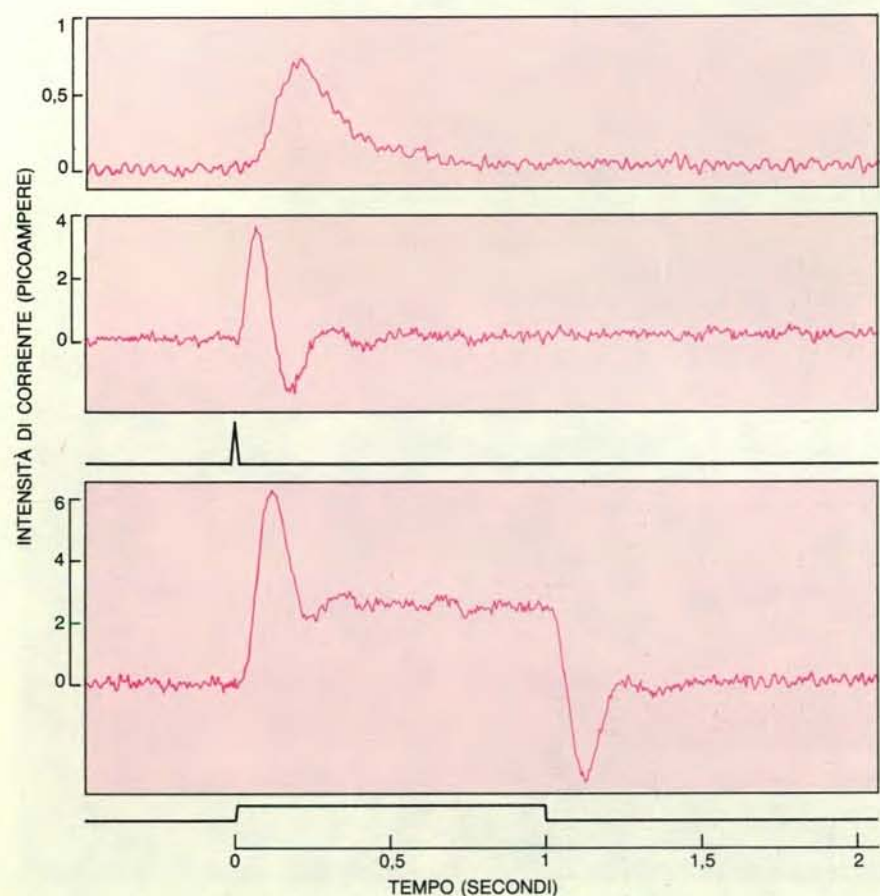
Per ovviare al problema posto da questo fenomeno, Yau, Trevor D. Lamb e uno degli autori (Baylor) decisero di usare un indicatore diverso dalla risposta dei bastoncelli. La scelta di misurare la fotocorrente nel bastoncello invece della sua differenza di potenziale si dimostrò felice perché la fotocorrente è effettivamente indipendente dal potenziale di membrana e quindi non viene influenzata dall'associazione dei segnali dei bastoncelli.

Per individuare la risposta a un singolo fotone, si osserva la risposta di un bastoncello a una luce molto debole. In un primo tempo abbiamo lavorato con bastoncelli ottenuti dalla retina di rospo; più recentemente, due di noi, in collaborazione con Brian J. Nunn di Stanford, hanno lavorato con coni e bastoncelli della retina di macaco (*Macaca fascicularis*). Per effettuare le misurazioni si introduce un singolo segmento esterno in un capillare di vetro e poi si registra la fotocorrente del bastoncello con un amplificatore molto sensibile collegato al capillare.

Sul bastoncello viene proiettato ripetutamente un lampo luminoso tanto debole da attivare in media una sola molecola di rodopsina. Si è visto che la fotocorrente risultante variava, assumendo valori prossimi a zero, uno, due e tre picoampere (10^{-12} ampere). Questa variazione è dovuta alle fluttuazioni casuali che si verificano nell'emissione dei fotoni da parte della sorgente luminosa, per cui talvolta il flash non attiva alcuna molecola di rodopsina, altre volte ne attiva una, due o tre. L'analisi statistica e la taratura dell'intensità del flash mostrano che la risposta di un picoampere è dovuta all'attivazione di una singola molecola di rodopsina. L'ampiezza e la forma della risposta si mantengono notevolmente costanti, suggerendo che esista un raffinato meccanismo di controllo enzimatico dell'amplificazione. Anche altre misurazioni indicano che c'è circa il 50 per cento di probabilità che un fotone produca, una volta assorbito, una risposta.

La capacità di rilevamento dei fotoni da parte dei bastoncelli è notevole, ma non perfetta. Anche in condizioni di buio completo, i bastoncelli producono occasionalmente un segnale identico a quello determinato dall'assorbimento di un fotone. Un bastoncello di retina di scimmia, per esempio, emette a caso segnali con la frequenza di uno ogni due minuti e mezzo in media. Sembra che questi segnali siano dovuti all'energia termica del tessuto che è in grado di attivare una molecola di rodopsina proprio come la luce. Questo fenomeno stabilisce il limite estremo della capacità di un bastoncello di rispondere in modo attendibile a una luce molto debole. Fortunatamente però l'attivazione termica procede abbastanza lentamente: considerando la frequenza dei segnali e il numero di molecole di rodopsina presenti nel bastoncello, possiamo calcolare che l'emivita del processo sia di 420 anni alla temperatura corporea. (È possibile studiare sperimentalmente i segnali erronei solo perché un bastoncello contiene una grandissima quantità di molecole di rodopsina.) Ciò nonostante, i segnali sono recepiti dal sistema visivo e provocano una sensazione di luce molto debole in condizioni di buio completo. Alcuni ricercatori che lavoravano nel campo della psicofisica hanno studiato quantitativamente tale fenomeno e hanno trovato, per il tasso di attivazione, valori simili a quelli calcolati da noi.

La risposta di un cono a un singolo fotone non può essere misurata perché è troppo piccola: le fluttuazioni di fondo la sovrastano. Nonostante ciò, la risposta a un singolo quanto di luce può essere valutata a partire dalla risposta di un cono ai lampi luminosi che attivano molte delle sue molecole di pigmento. Abbiamo calcolato che un solo fotone assorbito da un cono produca una fotocorrente la cui intensità è di circa 10 femtoampere (10×10^{-15} ampere), cioè una risposta circa cento volte inferiore rispetto a un bastoncello. La differenza di risposta delle due cellule fotorecettrici ci aiuta a capire perché la visione umana alla luce del giorno mediata dai coni è meno sensibile della visione notturna mediata dai bastoncelli.

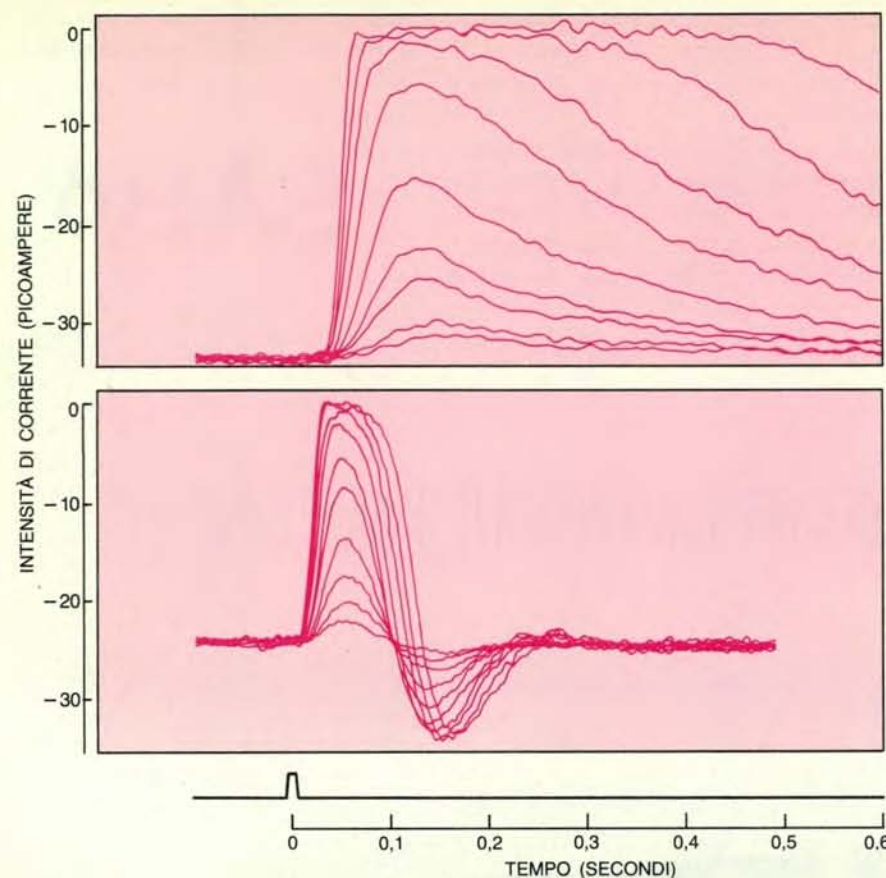
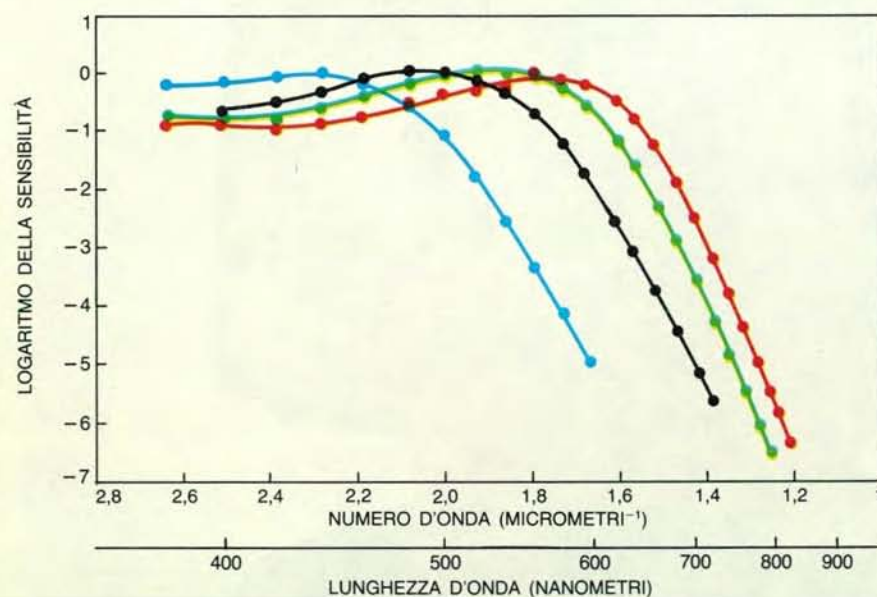
D'altra parte la risposta di un cono è approssimativamente quattro volte più veloce di quella di un bastoncello. Un bastoncello di primate impiega, per esempio, 300 millisecondi per condurre a termine il processo di emissione di un segnale in seguito all'assorbimento di un fotone. Grazie alla loro maggiore ve-

Il segmento esterno di un singolo bastoncello viene aspirato dentro una micropipetta per registrarne le correnti elettriche indotte dalla luce. Il segmento esterno nella pipetta è lungo circa 50 micrometri (10^{-6} metri) e ha un diametro di sei micrometri. Il bastoncello viene stimolato da un raggio di luce trasversale e la risposta elettrica viene quindi amplificata e registrata. Per evitare l'adattamento della cellula al buio, l'esperimento è condotto in luce infrarossa ed è seguito su uno schermo video. Il pezzetto di retina è stato estratto da un rospo, *Bufo marinus*.

La risposta di un bastoncello di scimmia a un singolo fotone è misurata introducendo il fotorecettore in una micropipetta. Nei due tracciati superiori il segmento esterno era colpito da flash deboli (*tratti più spessi*) che attivavano in media una molecola di pigmento. La risposta del bastoncello aveva un'ampiezza che variava da zero a uno o due picoampere (10^{-12} ampere). L'attivazione di una sola molecola di pigmento determina una risposta di un picoampere circa. Le fluttuazioni di ampiezza sono determinate da variazioni casuali dell'emissione di fotoni da parte della sorgente di luce. Il tracciato inferiore mostra la risposta dello stesso bastoncello a una luce debole, ma costante, che attivava in media circa 10 molecole di pigmento al secondo.

La forma d'onda dell'effetto prodotto da un singolo fotone in un fotorecettore di macaco è ricostruita mediando risposte a flash deboli. La variazione di corrente transmembrana è in funzione del tempo trascorso dal flash. La risposta del bastoncello (*in alto*) era indotta da un flash che attivava in media una molecola di pigmento; la risposta del cono (*al centro*) era provocata da un flash attivante circa 200 molecole. In basso è mostrata la risposta del cono a un impulso di un secondo.

locità di risposta, i coni riescono a percepire meglio stimoli visivi in rapido cambiamento.

Sembra quindi che nella trasduzione visiva sensibilità e risoluzione temporale si compensino: le piccole, veloci risposte dei coni permettono al sistema visivo di rilevare improvvisi cambiamenti di intensità o movimenti rapidi di oggetti, quando il livello di illuminazione è elevato e i bastoncelli sono saturati; d'altra parte, i segnali più lenti e più intensi di questi ultimi sono i più adatti per rilevare i fotoni, quando il livello di illuminazione è basso.

In condizioni di bassa intensità luminosa si assiste a uno straordinario incremento della sensibilità visiva grazie al passaggio dalla visione mediata dai coni a quella mediata dai bastoncelli. Entrando in una stanza debolmente illuminata, per esempio, diventiamo temporaneamente ciechi, a causa dell'insensibilità del sistema dei coni; lentamente però il sistema dei bastoncelli diviene più sensibile e, a mano a mano che esso viene assumendo il ruolo primario nel processo visivo, gli oggetti ritornano visibili. Comunque, anche nel caso della visione dovuta esclusivamente ai bastoncelli, la sensibilità visiva aumenta con il diminuire dell'intensità della luce di fondo. Questo cambiamento della sensibilità avviene all'interno dei bastoncelli o in altri neuroni che elaborano i segnali provenienti dai bastoncelli?

L'effetto della luminosità di fondo sulla sensibilità del bastoncello di un primate può essere determinata registrando le risposte a deboli lampi luminosi di un bastoncello adattato al buio. Dividendo l'ampiezza del picco di risposta per l'intensità del flash si ottiene una misura chiamata sensibilità al flash. Si accendono quindi le luci di fondo fisse e si determina ancora la sensibilità: questa diminuisce al crescere dell'intensità delle luci di fondo. Si può spiegare questa perdita di sensibilità con un semplice meccanismo di saturazione: al crescere dell'illuminazione di fondo si osserva la chiusura di un maggior numero di canali del sodio presenti nella membrana superficiale e questo fenomeno fa sì che diminuisca il numero di canali che possono reagire al flash.

Nel corso di esperimenti psicofisici si è calcolata la sensibilità del sistema dei bastoncelli umani, determinando quale sia l'intensità minima di un flash rilevabile in presenza di una luce di fondo diffusa. La sensibilità al flash diminuisce al crescere dell'intensità della luce di fondo. Una luce di fondo che desensibilizzi notevolmente la visione mediata dai bastoncelli può avere comunque uno scarso effetto sulla sensibilità misurata su un singolo bastoncello: per esempio, una luce regolare che attivi 40 molecole di rodopsina al secondo per bastoncello fa diminuire di 10 000 volte la sensibilità della visione mediata dai bastoncelli, ma

Le correnti di membrana sono state registrate in un bastoncello (*in alto*) e in un cono (*in basso*) di scimmia, illuminando uniformemente i segmenti esterni con flash di luce. Le registrazioni sovrapposte riportano la corrente nel segmento esterno in funzione del tempo trascorso dall'emissione del flash. L'intensità dei flash veniva progressivamente raddoppiata finché la risposta aveva ampiezza massima e la corrente verso l'interno si interrompeva del tutto. Nei bastoncelli un'ampiezza pari a metà del massimo si aveva quando venivano attivate 30 molecole di rodopsina; nei coni la stessa risposta era prodotta quando venivano attivate 1200 molecole di pigmento.

L'andamento relativo della sensibilità a un fotone dei bastoncelli e dei coni del macaco in funzione della lunghezza d'onda del fotone indica che la sensibilità spettrale è abbastanza simile a quella dei recettori umani. La curva in nero rappresenta lo spettro dei bastoncelli, mentre le curve in rosso, verde e blu corrispondono agli spettri di assorbimento dei coni per il rosso, il verde e il blu.

riduce la sensibilità di un segmento esterno di bastoncello solo del 20 per cento circa. Queste misurazioni hanno fornito un supporto sperimentale alle ipotesi del compianto William A. H. Rushton dell'Università di Cambridge secondo le quali, nell'uomo, la perdita di sensibilità da parte dei bastoncelli indotta dalla luce di fondo è dovuta all'attività dei neuroni che si trovano oltre il segmento esterno. Comunque il meccanismo che produrrebbe questo effetto deve essere ancora chiarito.

Quando l'intensità della luce di fondo supera il livello corrispondente a quello della luce a mezzogiorno, la sensibilità al flash della visione umana mediata dai bastoncelli diminuisce drasticamente a causa di un fenomeno di saturazione. In queste condizioni di luce forte il sistema visivo non rivela cambiamenti nella velocità con cui i bastoncelli assorbono fotoni. Misurazioni elettriche mostrano che singoli bastoncelli diventano incapaci di rispondere a un flash con intensità uguale all'incirca a quella di fondo, un'intensità cioè che satura la visione dei bastoncelli stessi. Questo limite nella visione caratterizza quindi la trasduzione effettuata dai bastoncelli.

La sensibilità di un singolo fotorecettore a luce di differenti lunghezze d'onda è determinata dalla probabilità con cui i suoi pigmenti assorbiranno fotoni di quella lunghezza d'onda. Misurazioni effettuate sulla sensibilità spettrale di singole cellule di macaco, che si ritiene abbiano fotorecettori simili a quelli umani, forniscono una base fisiologica per la spiegazione delle caratteristiche spettrali della visione nell'uomo. Le lunghezze d'onda della luce visibile si collocano approssimativamente tra i 400 e i 750 nanometri (10^{-9} metri). Luce di lunghezza d'onda maggiore (prossima all'infrarosso) è scarsamente assorbita dai pigmenti visivi; luce di lunghezza d'onda inferiore (prossima all'ultravioletto) può essere assorbita dai pigmenti visivi, ma non raggiunge la retina perché viene assorbita dalla cornea e dal cristallino.

I bastoncelli dei macachi mostrano un picco di sensibilità nella regione blu-verde dello spettro, prossima ai 490 nanometri. La sensibilità spettrale misurata concorda con quella della visione umana dei bastoncelli determinata in esperimenti psicofisici.

I coni nella retina delle scimmie si ripartiscono in tre gruppi che mostrano picchi di assorbimento prossimi a 430, 530 e 560 nanometri. I gruppi, che corrispondono strettamente a quelli umani, possono essere chiamati blu, verde e rosso per indicare le posizioni relative dei massimi nello spettro. Ciascun tipo di cono è sensibile a radiazioni luminose in un vasto ambito di lunghezze d'onda e le sensibilità dei vari gruppi mostrano una notevole sovrapposizione. Nonostante questo, sembra che la ripartizione dei pigmenti nei coni specifici sia relativa-

mente ristretta. In particolare, considerando la forma delle curve di sensibilità, possiamo concludere che meno di una molecola di pigmento su 100 000 in un cono della regione del blu risponde al rosso o al verde.

La risposta di un cono è indipendente dalla lunghezza d'onda del fotone che ha assorbito: tutti gli stimoli che provocano assorbimenti identici danno risposte identiche. Valutando il rapporto di eccitazione dei tre tipi di coni, il sistema visivo è comunque in grado di stimare il colore dalla lunghezza d'onda. Già da molti anni gli esperimenti psicofisici hanno chiarito che due stimoli composti da lunghezze d'onda differenti appariranno identici se entrambi gli stimoli provocheranno, all'interno di ciascun tipo di cono, lo stesso numero di assorbimenti. Sebbene tali osservazioni abbiano chiarito che la visione policromatica è fondata su tre colori, le basi esatte di questo meccanismo sono rimaste oscure a causa dell'incertezza riguardante le sensibilità specifiche alle varie regioni dello spettro per i diversi coni. Queste sensibilità oggi sono note per i coni della retina delle scimmie ed è confortante pensare che esse ci permettano di prevedere le regole che stabiliscono le intensità di luce richieste dall'uomo per realizzare accordi di colori.

In corrispondenza di lunghezze d'onda elevate, la percezione del colore è determinata soltanto dal relativo assorbimento dei coni rossi e verdi. A lunghezze d'onda superiori, al di là dei 600 nanometri circa, la tinta percepita cambia dall'arancio a un rosso progressivamente più intenso. Oltre i 700 nanometri interviene una curiosa inversione e la tinta diventa più arancio. Questo fenomeno, «lo slittamento paradossale del colore» (*paradoxical hue shift*), che venne scoperto nel 1955 da Giles S. Brindley di Cambridge, è spiegato dall'andamento della sensibilità allo spettro dei coni del rosso e del verde: il rapporto delle sensibilità di questi coni ha un massimo a 700 nanometri per cui questa lunghezza d'onda appare più rossa.

Ci vorrà ancora molto tempo per capire completamente il meccanismo molecolare che sta alla base della trasduzione visiva e il processo di elaborazione centrale dei segnali prodotti dai fotorecettori. Sebbene il mediatore interno per l'eccitazione visiva sia stato identificato, rimangono ancora da chiarire il meccanismo con cui opera la reazione a cascata dei nucleotidi e il sistema di controllo della permeabilità al sodio. Inoltre ci sono ancora molti punti oscuri per quanto riguarda le modalità con le quali i piccoli segnali generati da singoli fotoni vengono trasmessi attraverso le sinapsi, distinti dal rumore di fondo e quindi elaborati dal sistema visivo. Gli anni a venire saranno ricchi di interessanti prospettive per quanto riguarda gli esperimenti in tutti e due questi campi di ricerca.

Terapia antivirale

Nuovi farmaci che sfruttano le tenui differenze esistenti a livello molecolare tra virus e cellula ospite riescono a distruggere il virus senza ledere la cellula entro cui esso è fisicamente e funzionalmente incorporato

di Martin S. Hirsch e Joan C. Kaplan

Nel 1941, l'introduzione della penicillina ha rivoluzionato la terapia, mettendo a disposizione dei medici un'arma versatile e potente per combattere le infezioni batteriche. A più di un quarantennio di distanza non esiste un'analoga panacea per le infezioni virali. Paradossalmente, ciò dipende dalla semplicità stessa del virus. Un farmaco che uccida un agente patogeno senza intossicare il paziente deve essere in grado di distinguere l'invasore dalle cellule dell'ospite. Poiché i batteri sono organismi relativamente complessi, si differenziano dalle cellule dei mammiferi per molti aspetti metabolici e possono riprodursi anche al di fuori delle cellule ospiti. Questo semplifica il problema di trovare sostanze che attacchino i batteri in modo specifico.

Al contrario, la semplicità dei virus è tale che essi possono replicarsi solo penetrando fisicamente all'interno di una cellula e impadronendosi dei suoi meccanismi biochimici in modo da farle produrre il proprio materiale genetico e le proprie proteine. Poiché il ciclo di replicazione dei virus è intimamente connesso con le funzioni delle cellule ospiti, i virus possiedono un minor numero di caratteristiche biochimiche adatte a un attacco selettivo tanto che, fino agli inizi degli anni sessanta, il problema sembrava addirittura insormontabile. La principale strategia per controllare le infezioni virali era (e lo è ancora in grande misura) la produzione di vaccini, che non attaccano direttamente un virus, ma ne prevengono l'infezione stimolando in anticipo il sistema immunitario.

Tuttavia, negli ultimi due decenni l'accumularsi delle conoscenze sulla replicazione virale ha permesso di definire eventi che sono specifici per i virus e di identificare composti che potrebbero interferire con essi. Prove di laboratorio, seguite da esami clinici accuratamente condotti, hanno confermato le promesse di alcuni di tali composti. Questi nuovi farmaci antivirali hanno già trasformato la terapia di diverse malattie. Meno di dieci anni fa un neonato sofferente per un'infezione diffusa da *Herpes simplex* o un adulto con encefalite erpetica aveva solo un 20-30 per cento di probabilità di sopravvivenza; oggi un trattamento farmacologico precoce e aggressivo di entrambe le malattie può aumentare questa probabilità a più del doppio. Ricercatori e clinici stanno ora cercando di applicare l'esperienza maturata negli ultimi 20 anni per fare fronte a un nuova sfida: individuare terapie per l'AIDS, la mortale malattia virale identificata per la prima volta nel 1981.

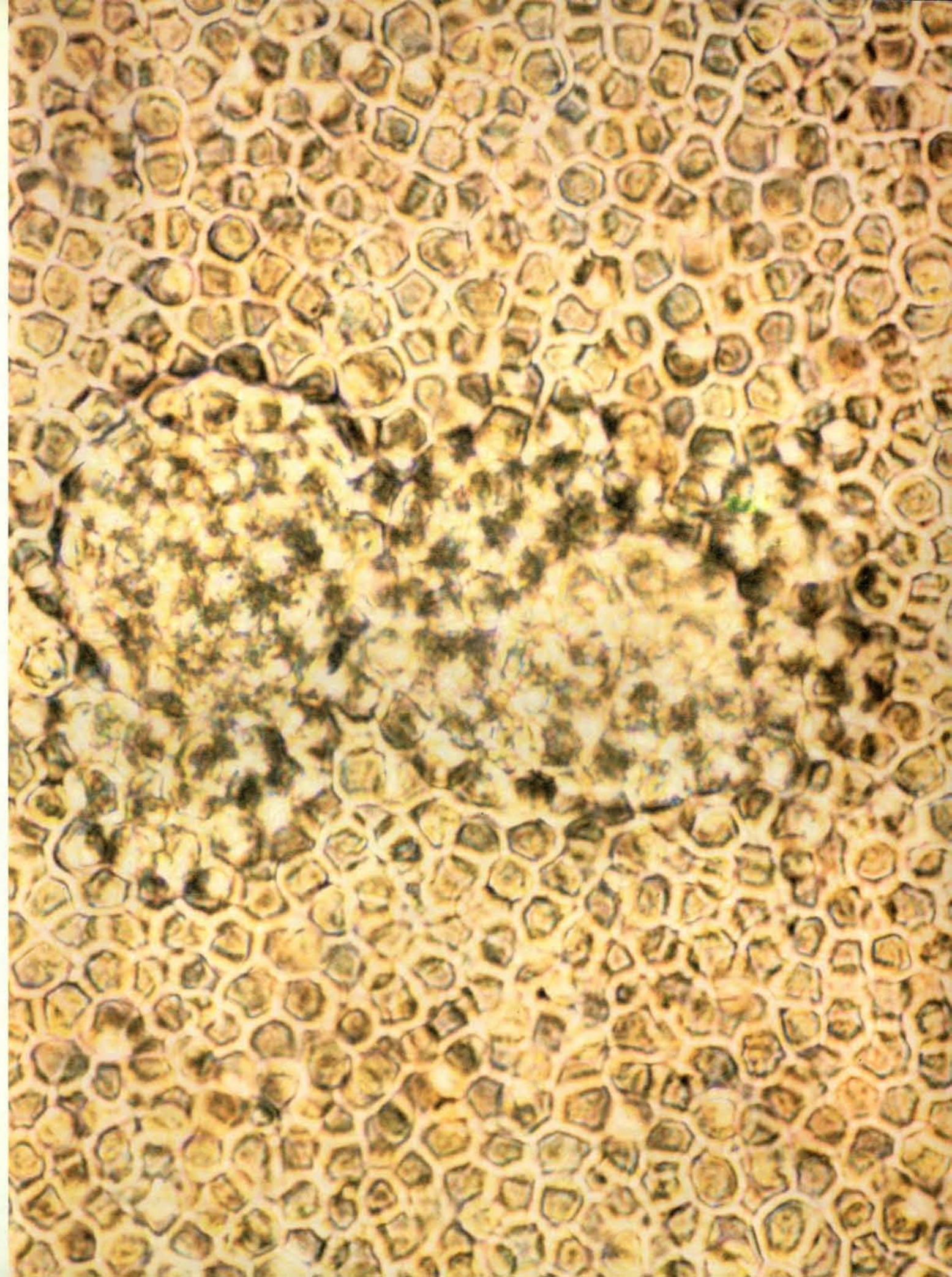
Modalità d'infezione

I tentativi per curare le infezioni virali devono tener conto delle varie modalità con cui i virus interagiscono con le cellule ospiti. Quando un virus infetta una cellula, si lega alla sua membrana superficiale, penetra nella cellula e si libera del proprio involucro proteico. Il materiale genetico, che può essere costituito da una molecola a filamento singolo o duplice sia di DNA sia di RNA, contiene i geni che specificano i pochi enzimi e le poche proteine strutturali di cui il virus necessita per interagire con la cellula e riprodursi. I meccanismi biochimici dell'ospite portano questi geni a esprimersi nelle fasi più tardive del ciclo di replicazione virale.

Queste fasi tardive determinano il successivo tipo d'infezione. In un'infezione litica, caratteristica di molti virus a DNA e a RNA comprendenti tra l'altro i virus del raffreddore e della poliomielite, il virus si replica inducendo la cellula a copiare il proprio materiale genetico e a produrre proteine virali. La proliferazione di nuove particelle virali fa sì che la cellula infetta si lisi, cioè scoppi liberando le particelle. Le infezioni litiche spesso si diffondono rapidamente in tutta la popolazione di cellule vulnerabili, distruggendole in una fase precoce della malattia. Pertanto, per essere efficace, il trattamento della maggior parte delle forme litiche deve cominciare precocemente e richiede sia una pronta diagnosi, sia una valida terapia.

La replicazione del virus in un altro tipo di infezione, l'infezione persistente, non uccide sempre la cellula infettata: le particelle virali vengono spesso liberate gradualmente e la cellula sopravvive intatta e continua a dividersi, anche se la sua funzione può subire insidiosi cambiamenti. La persistenza può anche assumere un'altra forma, in cui alcune cellule ospitano un'infezione litica, mentre altre non sono interessate. Nell'uno e nell'altro caso l'infezione può continuare per mesi o anni senza provocare una sintomatologia manifesta. Per esempio, l'epatite *B* può dar luogo a infezioni che persistono indefinitamente, dando origine solo in alcuni casi a una malattia epatica come un cancro o una vasculite. Anche i virus della leucemia umana e dell'immunodeficienza umana (HIV), agente etiologico dell'AIDS, possono produrre sintomi solo dopo un'infezione che duri già da molto tempo. (HIV è la sigla oggi accettata per il virus che, in origine, era chiamato HTLV-III o LAV.)

In entrambe le infezioni, litica e per-

In una coltura infettata con il virus dell'immunodeficienza umana (HIV), l'agente etiologico dell'AIDS, la replicazione del virus ha danneggiato i linfociti *T* umani, le cellule che sono gli ospiti di elezione dell'HIV. I componenti dell'involucro virale, sintetizzati allorché il virus si è moltiplicato, hanno alterato la membrana superficiale di queste cellule immunitarie dando origine ad aggregati multinucleati (sincizi). Nelle prove di laboratorio eseguite sui farmaci antivirali messi a punto per la terapia dell'AIDS, la capacità di impedire che i sincizi si sviluppino fino a dare origine a una coltura infetta è considerata una misura dell'efficacia di questi prodotti.

sistente, il virus completa il proprio ciclo di replicazione con un processo le cui numerose, singole fasi potrebbero venir interrotte dal farmaco antivirale. Un terzo tipo d'infezione, l'infezione latente, non offre simili opportunità: un virus latente non si riproduce e mancano le attività biochimiche associate alla riproduzione. Un'infezione di questo tipo è davvero difficile da individuare, poiché non è contrassegnata dalla morte delle cellule infettate o dalla produzione di nuove particelle virali. Il materiale genetico del virus può integrarsi nei cromosomi della cellula cosicché, durante la divisione cellulare, i geni virali si riproducono insieme a quelli cellulari e vengono trasmessi alle cellule figlie. In certe condizioni i virus possono essere riattivati con il conseguente sviluppo di un'infezione attiva.

La latenza, che si alterna a periodi di replicazione virale attiva, caratterizza molti virus erpetici: l'*Herpes simplex* (agente etiologico di una gamma di infezioni tra cui l'herpes labialis e genitalis), il virus della varicella-*zoster* (agente etiologico della varicella e del fuoco di Sant'Antonio), il citomegalovirus e il virus di Epstein-Barr. I farmaci antivirali prodotti finora possono influire su questi virus solo durante la fase di replicazione. L'attacco a un virus nel periodo di latenza comporterebbe la possibilità di distinguere il materiale genetico virale da quello dell'ospite e la sua distruzione selettiva. Un simile evento presuppone addirittura nuovi strumenti concettuali.

La preparazione di un nuovo farmaco

La conoscenza del ciclo di replicazione di un virus e della sua interazione con le cellule ospiti può fornire indicazioni su quali potrebbero essere le funzioni specifiche del virus vulnerabili all'attacco da parte di un farmaco. Il percorso dalla prima fase concettuale alla produzione di un farmaco antivirale efficace è lungo e dispendioso. Per trovare una sostanza che inibisca una funzione virale può rivelarsi necessario l'esame in laboratorio di centinaia di composti. Se un composto rivela qualche attività antivirale, si compiono ulteriori prove *in vitro* su di esso così da valutarne l'efficacia in modo più preciso e determinarne la tossicità.

Il composto viene sperimentato su una varietà di cellule infettate dal virus, tra le quali anche cellule di origine umana, e per concentrazioni differenti sia del farmaco sia del virus. Le misurazioni che riguardano la produzione del virus, i saggi sulle proteine virali e le osservazioni del danno cellulare (come la lisi) provocato dalla replicazione virale servono a dare indicazioni sull'efficacia del farmaco. Nel contempo, le cellule che ospitano il virus sono tenute sotto osservazione per rilevare segni di tossicità del farmaco. Fra i preparati virali contro i quali si prova il composto dovrebbe esserci altrettanto materiale isolato da pazienti infetti quanto ceppi cellulari di laboratorio, che in molti casi dopo anni di coltura ininterrotta potrebbero aver accumulato mutazioni genetiche tali da alterarne la reazione al farmaco.

Se le promesse iniziali del composto sono confermate in provetta, occorre procedere studiando la farmacocinetica del composto - ossia, come l'organismo lo trasporta, lo metabolizza e lo elimina - somministrandolo in primo luogo ad animali di laboratorio. La capacità di certi virus, tra i quali anche l'HIV, di colpire il cervello e il midollo spinale creandovi una riserva di virus potenzialmente infettivi ha un certo rilievo per alcuni studi condotti su animali. I capillari cerebrali hanno la particolarità di creare una barriera ematoencefalica, che impedisce a molti composti di entrare nel sistema nervoso centrale. Gli studi su animali consentono di osservare la capacità di un farmaco di attraversare tale barriera e danno inoltre informazioni sulla tossicità del farmaco. Spesso, tuttavia, le prove su animali non possono dimostrare l'efficacia di un farmaco: molti virus umani o non infettano altre specie o, se le infettano, non sono patogeni.

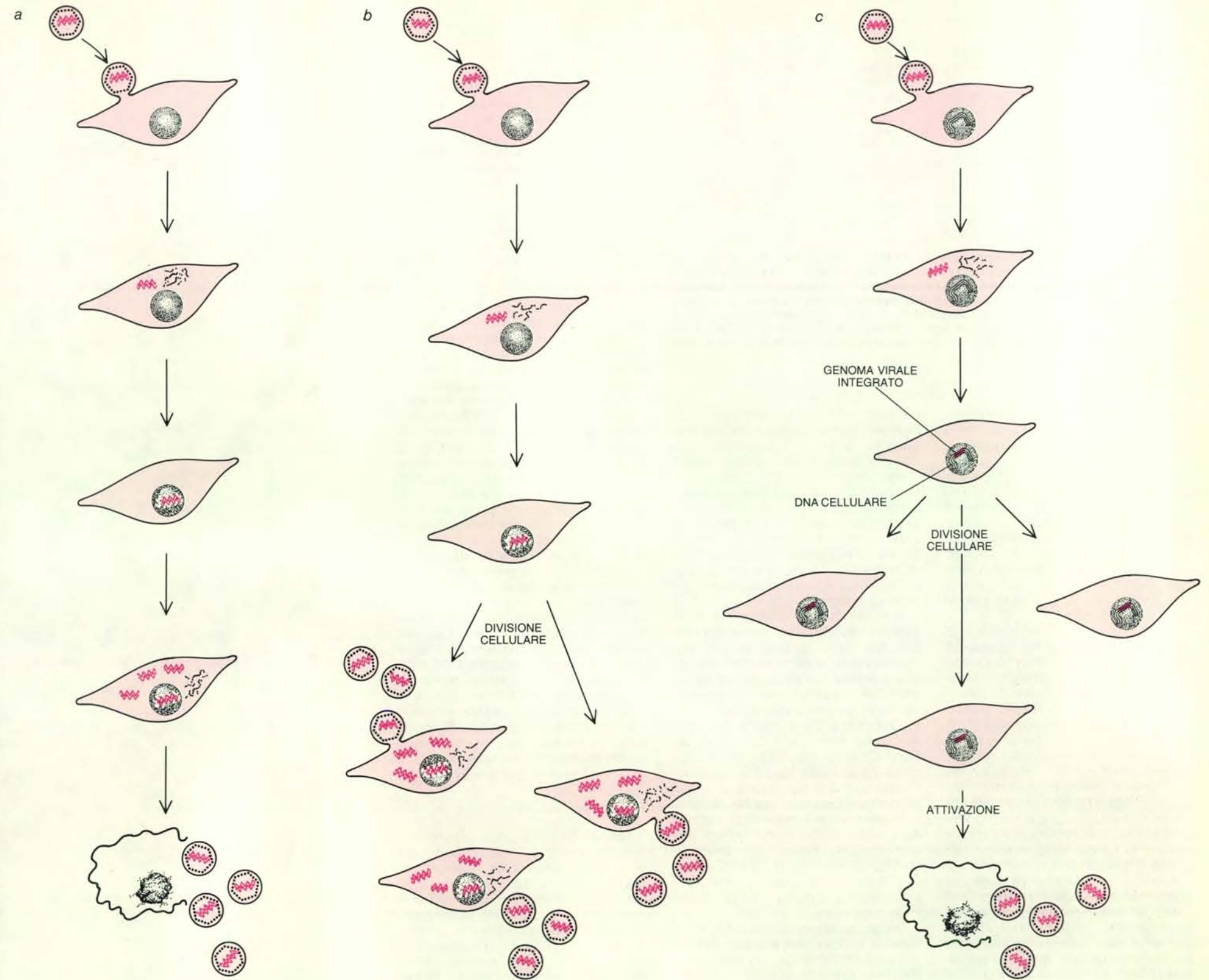
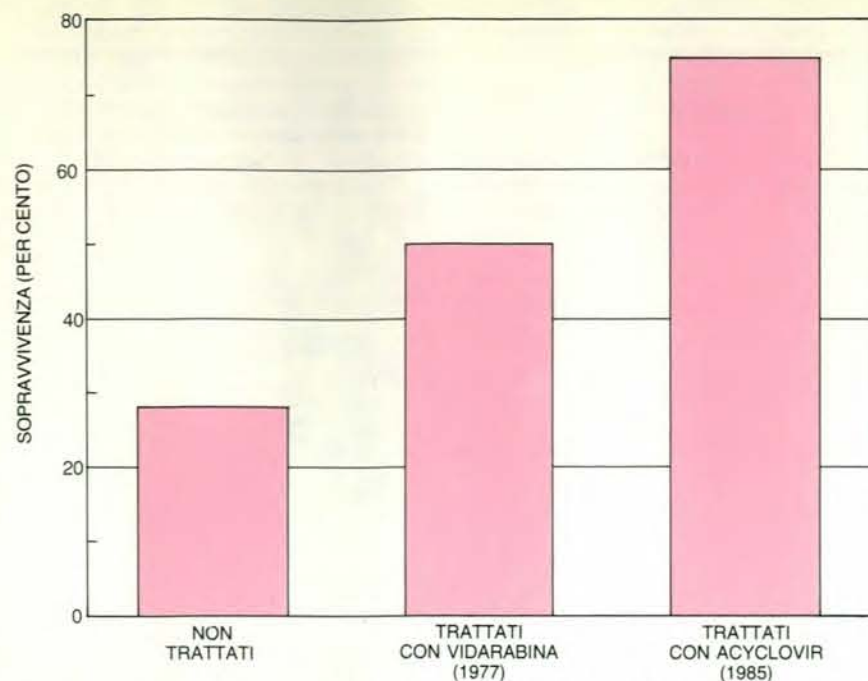
Questi studi sono seguiti da una fase di prove chimiche preliminari su piccoli gruppi di pazienti, che fornisce ulteriori informazioni sulla farmacocinetica del prodotto e sulla sua sicurezza. In particolare, esse servono a dare un'indicazione sulla dose massima tollerata.

Le prove cliniche su vasta scala rappresentano la fase finale del protocollo eseguito per valutare la sicurezza e l'efficacia del farmaco; esse devono essere effettuate su una popolazione di pazienti accuratamente selezionata. Per esempio, è spesso importante che i pazienti trattati si trovino a uno stesso stadio della malattia. Poiché la replicazione del virus può raggiungere il massimo subito dopo l'insorgenza delle manifestazioni cliniche, il farmaco in esame deve essere provato su pazienti nei quali l'infezione sia stata diagnosticata di recente.

Poiché il decorso naturale della maggior parte delle infezioni virali è molto variabile, le prove devono essere rigorosamente oggettive e attentamente controllate. Se per l'infezione non esiste un trattamento certo, si usa naturalmente quale termine di confronto un placebo. La popolazione esaminata viene suddivisa, di solito a caso, in due o più gruppi, ad almeno uno dei quali è somministrato il farmaco, mentre a un altro, il gruppo di controllo, viene dato il placebo. In genere il metodo è a doppio cieco, in modo che né i pazienti né i ricercatori sappiano chi ha ricevuto il farmaco. Queste prove sono impegnative e hanno costi elevati, ma nell'ultimo decennio hanno permesso di stabilire in modo valido l'efficacia dell'attuale generazione di farmaci antivirali.

Un farmaco modello

Un importante successo conseguito in campo farmacologico è rappresentato dall'acyclovir, un composto brevettato di recente, che inibisce in modo molto specifico la replicazione di certi virus erpetici. Esso ha in gran parte soppiantato le precedenti generazioni di farmaci contro le malattie erpetiche che non sono né altrettanto efficaci, né specifici.

Che l'acyclovir potesse inibire una funzione virale, e precisamente la sintesi

A seconda del tipo, virus e cellule ospiti possono interagire in maniera diversa. Un'infezione litica (*a*) produce la morte della cellula infettata. Dopo che il virus, penetrato nella cellula, si libera del proprio involucro proteico, il suo materiale genetico sovverte il meccanismo biochimico cellulare producendo nuove proteine e nuovo materiale genetico che concorrono successivamente a formare una grande quantità di particelle virali. La cellula ospite muore allorché tali particelle sono riversate all'esterno. In un'infezione persistente (*b*), la replicazione virale procede, ma la cellula può sopravvivere e continuare a moltiplicarsi e a liberare virus. In un'infezione latente (*c*), la replicazione virale è invece sospesa. Il genoma virale può integrarsi nel corredo cromosomico della cellula per essere copiato e trasmesso alle cellule figlie nel corso della divisione cellulare. In determinate circostanze il virus latente può essere riattivato in alcune delle cellule infette.

La sopravvivenza in pazienti affetti da encefalite da *Herpes simplex*, un'infezione virale che colpisce il cervello, è aumentata in maniera molto marcata con l'avvento di farmaci efficaci e specifici per le infezioni erpetiche. Nel 1977 si è dimostrato che il trattamento con vidarabina aumentava la sopravvivenza come conseguenza della capacità del composto di interferire con la sintesi del DNA virale e quindi con la replicazione del virus. Nel 1985 è stato trovato che l'acyclovir aveva un effetto ancora più rilevante: alcuni aspetti della replicazione di *Herpes simplex* rendono il virus particolarmente vulnerabile all'acyclovir e a farmaci caratterizzati da struttura analoga.

di nuovo DNA nel corso della replicazione, è stato suggerito per la prima volta dalla struttura della sua molecola, che è simile a quella di un nucleoside, un precursore del DNA o dell'RNA. Un nucleoside consiste di una o l'altra di quattro basi (gruppi chimici ad anello singolo o doppio, noti rispettivamente come pirimidine e purine), legata a uno zucchero a cinque atomi di carbonio. Un analogo di nucleoside consiste, alla stessa stregua, di una base, o di un derivato di una base, e di uno zucchero, o di un gruppo simile a uno zucchero.

Nel corso normale della sintesi di DNA, i nucleosidotrifosfati (nucleosidi con tre gruppi fosfato) liberi reagiscono con l'estremità di una catena di DNA in fase di allungamento. Durante la reazione, che è mediata da un enzima, la DNA-polimerasi, un gruppo ossidrilico (OH) in posizione 3' dell'anello di zucchero che si trova in corrispondenza della estremità della catena forma un legame con il gruppo fosfato in posizione 5' del nucleosidotrifosfato che si aggiunge alla catena. Gli altri due gruppi fosfato vengono liberati e alla catena resta così legato un nucleotide, cioè un nucleoside con un singolo gruppo fosfato.

Allorché, nel corso degli anni cinquanta, si è giunti alla conoscenza della sintesi del DNA, è risultato evidente che un nucleoside modificato (cioè un analogo di nucleoside) avrebbe potuto interferire con il processo. Analoghi dei nucleosidi vennero all'inizio sintetizzati come potenziali farmaci antitumorali: si pensava che essi potessero rallentare o bloccare la produzione accelerata di DNA che ha luogo nelle cellule neoplastiche, che si dividono rapidamente. Quando nel corso degli anni sessanta si comprese che alcuni virus innescavano una rapida sintesi del DNA in cellule che avevano infettato, vennero provati per la loro attività antivirale molti analoghi di nucleosidi. L'acyclovir, un analogo della guanosina, era uno tra i primi, promettenti composti.

Si scoprì che l'acyclovir aveva un effetto antivirale altamente specifico e che la sintesi di DNA da parte di certi virus erpetici era sensibilissima a esso, mentre non si poteva dire altrettanto per la sintesi di DNA in cellule normali. Con una ricerca prolungatasi per una decina d'anni, Gertrude B. Elion e collaboratori dei Wellcome Research Laboratories nel North Carolina, hanno chiarito i meccanismi che sono alla base dell'azione antivirale specifica del farmaco. Essi scoprirono che per interferire con la sintesi del DNA virale una molecola di acyclovir deve acquistare tre gruppi fosfato, proprio come un nucleoside deve essere fosforilato tre volte prima di prendere parte alla sintesi del DNA cellulare. Ogni fosforilazione è mediata da un enzima noto come chinasi.

Il compito di una chinasi nella cellula consiste nel trasferire gruppi fosfato da una molecola all'altra. Le chinasi cellulari fanno da mediatori nella formazione dell'acyclovirdifosfato e dell'acyclovirtrifosfato, ma la fase iniziale, cioè la conversione dell'acyclovir in acyclovirmonofosfato, dipende in gran parte da una chinasi specificata da un gene del virus erpetico, la timidinchinasi, presente solo nelle cellule infette. Nelle cellule sane esiste un enzima che fosforila l'acyclovir, ma in misura molto limitata: la quantità di acyclovirtrifosfato che ne risulta è di vari ordini di grandezza inferiore alla quantità che si forma nelle cellule infette. Pertanto, il farmaco è attivo soprattutto nei confronti di queste ultime.

Nel danneggiare la sintesi del DNA, l'acyclovir identifica anche il DNA virale. Poiché il composto assomiglia a un vero nucleoside, una DNA-polimerasi può attaccarsi alla sua forma attivata (cioè acyclovirtrifosfato) e unire a un filamento di DNA in fase di allungamento un'unità di acyclovirmonofosfato, mentre gli altri due gruppi fosfato vengono liberati durante il processo. Per ragioni non ancora chiare, la DNA-polimerasi dei virus erpetici ha un'affinità per il farmaco molto superiore rispetto alla polimerasi cellulare; pertanto, le molecole di acyclovirmonofosfato vengono addizionate preferenzialmente al DNA virale piuttosto che al DNA cellulare. Quando il farmaco si attacca al filamento di DNA, forma un legame con un ossidrile in posizione 3' all'estremità del filamento, proprio come farebbe un nucleotide convenzionale. Diversamente da un nucleotide vero, tuttavia, l'acyclovirmonofosfato manca di un gruppo ossidrile in posizione 3' e non offre quindi alcun punto di attacco a un altro eventuale legame. Così, una volta che esso sia stato incorporato nel filamento di DNA, l'allungamento di quest'ultimo cessa.

Inoltre, mentre una DNA-polimerasi di solito si lega solo temporaneamente al filamento di DNA che si sta allungando, dimodoché ogni molecola dell'enzima può catalizzare illimitatamente l'allungamento del DNA, l'acyclovirmonofosfato si associa con il DNA in modo permanente. La polimerasi virale si lega irreversibilmente a quel nucleotide «fraudolento» che ha appena aggiunto al DNA. Cessano così, contemporaneamente, sia l'allungamento del filamento di DNA virale, sia l'utilità della polimerasi virale. La particolare affinità esibita da acyclovir verso gli enzimi virali si traduce così in una potente azione antivirale con il minimo di tossicità per la cellula ospite.

Gli usi dell'acyclovir

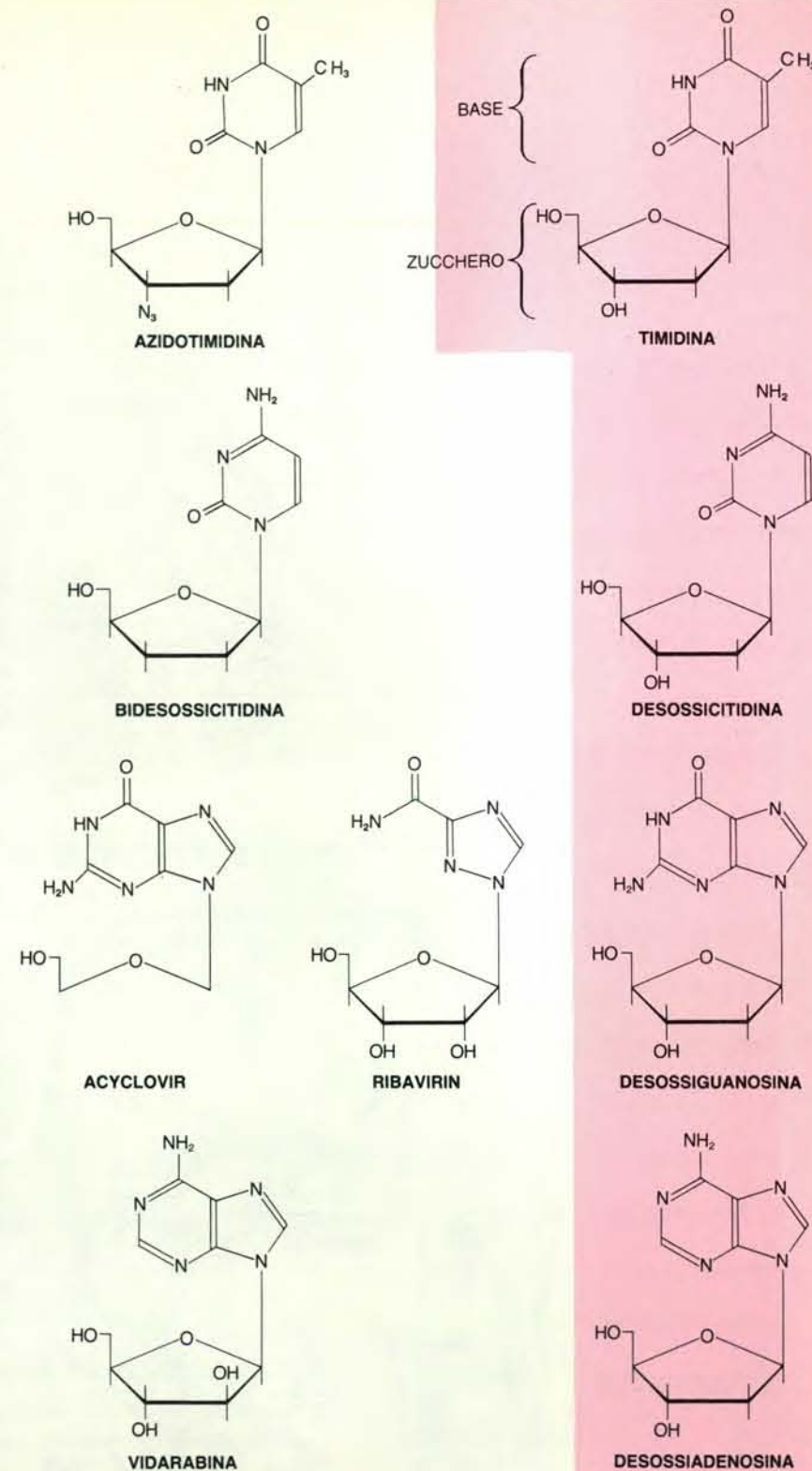
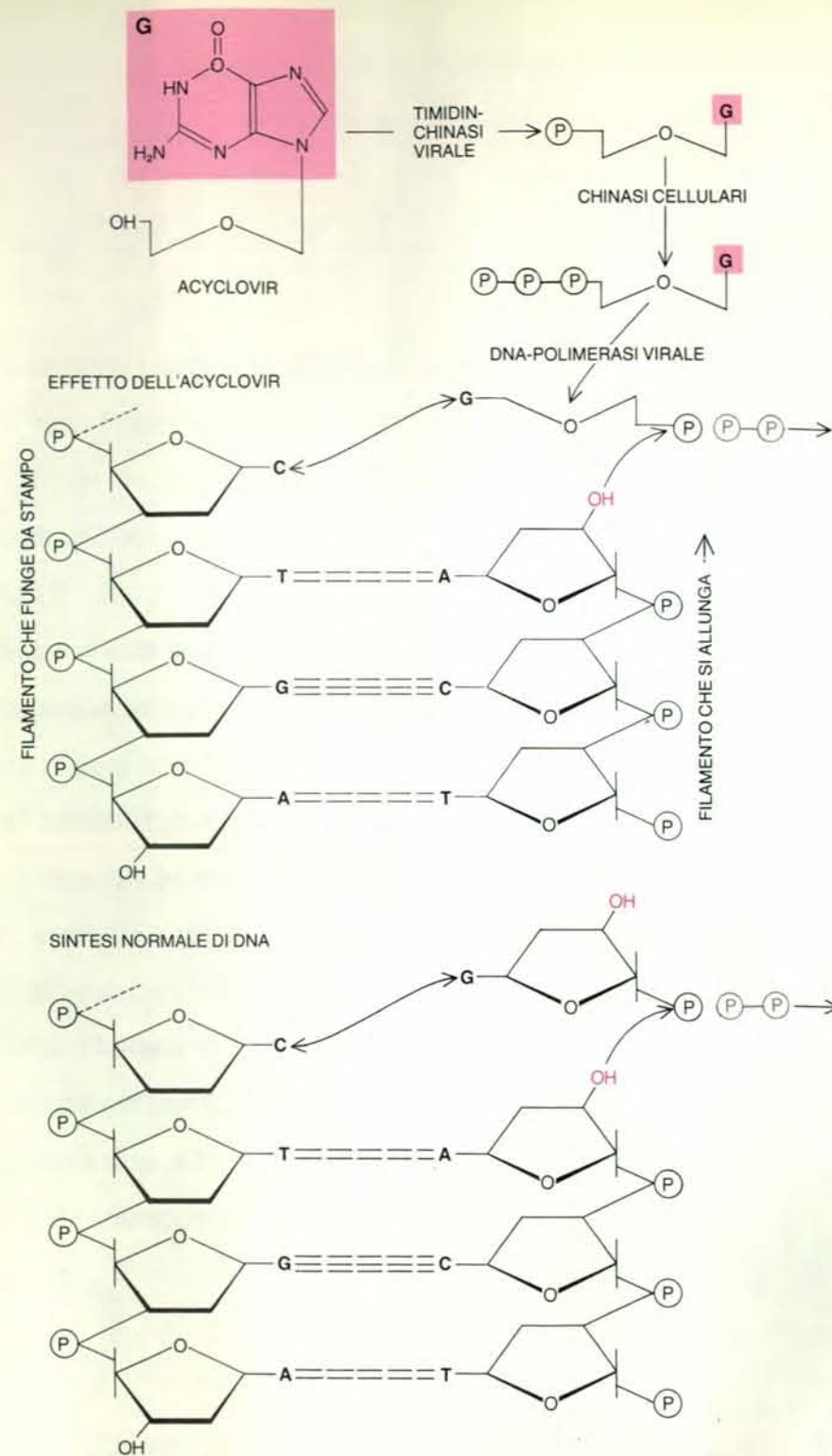
L'acyclovir trova la sua massima applicazione nel trattamento delle infezioni da *Herpes simplex* che colpiscono gli organi genitali e che interesserebbero, nei soli Stati Uniti, più di 20 milioni di persone. L'attacco iniziale, cioè il primo episodio di infezione litica, è generalmente il più grave e più prolungato e rappresenta la fase d'elezione per un trattamento con il farmaco. La terapia

con acyclovir (che può essere somministrato per via endovenosa, come pomata o per via orale) allevia i sintomi, riduce la quantità di virus liberato dalla zona infetta e accelera la guarigione. Essa non previene successivi attacchi e neppure ne diminuisce la frequenza o la gravità. Somministrazioni orali di acyclovir a scopo preventivo possono avere solo un modesto influsso sulle infezioni ricorrenti, perciò nella maggior parte dei casi il costo del farmaco e la sua potenziale tossicità sul lungo termine non giustificano un simile tipo di cura.

L'acyclovir si è dimostrato utile anche nella varicella e nel fuoco di Sant'Antonio provocati dal virus varicella-*zoster*. In pazienti in cui il sistema immunitario è stato depresso per poter procedere al trapianto di un organo o è stato danneggiato da una malattia, l'acyclovir può talvolta avere una funzione preventiva, evitando le infezioni da *Herpes simplex* e da varicella-*zoster*. Il farmaco può anche prevenire, in questi soggetti, la riacutizzazione di un'infezione erpetica latente. All'acyclovir va attribuito, infine, il merito della marcata diminuzione del tasso di mortalità per encefaliti da *Herpes simplex*, rilevata nell'ultimo decennio.

È possibile che anche per l'acyclovir si presenti lo stesso problema che si è dovuto affrontare nel caso dei farmaci antibatterici dopo un certo numero di anni dalla loro introduzione: la comparsa di agenti patogeni resistenti. Mutanti di virus erpetici resistenti all'acyclovir possono essere ottenuti in laboratorio coltivando i virus in presenza del farmaco. Ceppi resistenti sono stati isolati anche da pazienti. La resistenza all'acyclovir può avere diverse origini genetiche: una mutazione che inattivi il gene virale per la timidinchinasi oppure, in alternativa, una mutazione che produca una timidinchinasi alterata con un'affinità ridotta per l'acyclovir. In entrambi i casi, l'acyclovir non viene fosforilato nelle cellule infette e, pertanto, rimane inattivo. Negli animali sembra che i mutanti per la timidinchinasi siano meno patogeni dei ceppi non mutanti e anche meno capaci di provocare un'infezione latente.

La resistenza può anche insorgere a seguito di una mutazione a carico della DNA-polimerasi virale, che diminuisce la sua capacità di interagire con l'acyclovirtrifosfato. Studi effettuati in laboratorio suggeriscono che, contrariamente ai mutanti per la timidinchinasi, questi ceppi resistenti possono essere altrettanto patogeni del virus sensibile al farmaco. Questo risultato suggerisce che una somministrazione prolungata di acyclovir a scopo preventivo o la sua prescrizione per infezioni banali potrebbe favorire la comparsa di ceppi virali patogeni resistenti al farmaco.

Per questi mutanti e per alcuni virus erpetici come il citomegalovirus e il virus di Epstein-Barr, per i quali non è noto se possiedano il gene per la timidinchinasi e che appaiono relativamente insen-

Gli analoghi dei nucleosidi sono una classe di composti sintetici che devono il loro nome e la loro attività antivirale a una somiglianza con i nucleosidi, i precursori chimici del DNA e dell'RNA. Qui è mostrato un certo numero di composti, alcuni già di uso clinico in caso di infezioni virali, altri in corso di sperimentazione, con a fianco i nucleosidi corrispondenti (*in colore*). Un nucleoside consta di una base ad anello singolo o doppio, legata a uno zucchero a cinque atomi di carbonio; un analogo differisce da esso per alcuni aspetti più o meno rilevanti della base o dello zucchero. Un enzima che normalmente agisce su un nucleoside può legarsi anche al suo analogo nel corso della replicazione virale . Tuttavia, a causa delle differenze che esistono tra nucleoside e analogo, il legame con quest'ultimo può inattivare l'enzima, interrompendo così un processo molecolare che è essenziale per la replicazione. Certi analoghi di nucleosidi hanno altre caratteristiche che si aggiungono alla specificità e all'efficacia che dimostrano contro i virus.

L'effetto dell'acyclovir in un'infezione da virus erpetico consiste nell'inibire la sintesi del DNA virale. In primo luogo, l'acyclovir deve acquisire i tre gruppi fosfato (*P*). Per addizionare il primo fosfato (*in alto*), il farmaco dipende dalla timidinchinasi, un enzima che è specificato da alcuni di questi virus. Le chinasi cellulari aggiungono gli altri due fosfati. La DNA-polimerasi virale, l'enzima che catalizza la sintesi del DNA virale, si lega quindi all'acyclovir come se fosse un normale nucleosidetrifosfato (unità costruttiva del DNA) e lo attacca a un filamento di DNA che si sta allungando (*al centro*). Uno dei gruppi fosfato sulla molecola dell'acyclovirtrifosfato forma un legame con l'ossidrile in posizione 3′ dello zucchero che si trova all'estremità della catena di DNA; gli altri due gruppi fosfato sono, invece, liberati. Diversamente da un nucleoside normale (*in basso*), l'acyclovir manca dell'anello di zucchero e di un ossidrile in posizione 3′. Pertanto non può consentire un ulteriore allungamento del filamento di DNA. Inoltre, la DNA-polimerasi virale, che normalmente continuerebbe a catalizzare reazioni di addizione al filamento di DNA, forma un complesso con il DNA e il farmaco e quest'ultimo, pertanto, non è più in grado di agire.

sibili all'acyclovir, sono state proposte strategie alternative. Un analogo di nucleoside di recente sintesi, la diidrossipropossimetilguanina (DHPG), esibisce una notevole attività contro i suddetti virus, sia in provetta, sia nei pazienti, anche se non sono note con esattezza le sue modalità d'azione. La vidarabina, un composto che prima dell'avvento dell'acyclovir rappresentava il punto di forza della terapia contro le più gravi infezioni virali, può agire anche in assenza di timidinchinasi. Come l'acyclovir, la vidarabina è un analogo di nucleoside che viene attivato nella cellula da una triplice fosforilazione e che quindi interferisce con la sintesi del DNA virale.

Altri successi nella lotta antivirale

Le infezioni virali a carico dell'apparato respiratorio rimangono una fra le principali cause di malattia, ma la produzione di farmaci ha consentito parecchi progressi nella cura dell'influenza e delle malattie provocate dal virus sinciziale respiratorio, come la polmonite e la bronchiolite della prima infanzia. Due composti che si sono dimostrati utili contro l'influenza, l'amantadina e la rimantadina, sono stati ottenuti prima dello sviluppo delle recenti acquisizioni teoriche nel campo della terapia antivirale. La loro efficacia contro l'influenza *A* è stata scoperta all'inizio degli anni sessanta nel corso di un'indagine eseguita in laboratorio su un gran numero di composti. Ancora oggi la loro modalità d'azione non è chiara. Alcuni ricercatori hanno avanzato l'ipotesi che amantadina e rimantadina impediscano al virus dell'influenza *A* di liberarsi del proprio rivestimento proteico allorché infetta una cellula, oppure che possano interferire con i primi stadi della trascrizione del materiale genetico virale in RNA messaggero, il quale serve da stampo per la sintesi proteica. Solo l'amantadina è brevettata negli Stati Uniti; essa per lo più viene somministrata preventivamente durante le epidemie di influenza.

Gli sforzi più recenti compiuti per bloccare fasi specifiche della replicazione virale mediante analoghi dei nucleosidi hanno portato alla produzione di ribavirin, un farmaco efficace contro il virus sinciziale respiratorio. Si tratta di un analogo della guanosina che, attivo *in vitro* contro una vasta gamma di virus, ha mostrato la propria efficacia anche in prove cliniche su pazienti affetti da febbre di Lassa, una malattia virale endemica in Africa.

Si ritiene che l'efficacia del ribavirin sia collegata ad almeno due funzioni biochimiche. Nella forma fosforilata, il farmaco interferisce con un enzima fondamentale per la sintesi dei precursori della guanosina del DNA e dell'RNA. La sua rassomiglianza con la guanosina gli permette anche di inibire l'«incappucciamento» (*capping*) dell'RNA messaggero virale, in cui una molecola di guanosina modificata viene aggiunta all'estremità dell'RNA messaggero appena trascritto. Sembra che il cappuccio permetta a questo RNA di unirsi ai ribosomi, che sono le strutture cellulari su cui vengono sintetizzate le catene polipeptidiche. Il processo di incappucciamento è universale, ma la molecola di ribavirin agisce selettivamente nei confronti dell'inibizione. Essa evita l'incappucciamento dell'RNA messaggero virale - e pertanto la produzione delle proteine virali - molto più di quanto interferisca con le corrispondenti funzioni delle cellule umane.

Anche così, tuttavia, il ribavirin può essere tossico per le cellule del sangue e può provocare anemia se somministrato per via orale. Recenti esperimenti hanno però dimostrato che esso può agire contro il virus sinciziale respiratorio anche quando viene inalato sotto forma di aerosol, una tecnica che fa pervenire il farmaco direttamente ai tessuti infettati, riducendone al minimo l'assorbimento e, pertanto, gli effetti tossici.

Gli interferon

La terapia antivirale ha compiuto progressi non solo sul fronte dei farmaci che inibiscono la replicazione virale, ma anche appropriandosi di una delle difese naturali dell'organismo, ossia della classe di proteine note come interferon. Il nome è stato coniato alla fine degli anni cinquanta quando si è scoperto che le cellule esposte a un virus secernono proteine che aumentano la resistenza a quell'infezione. Tuttavia gli interferon sono stati purificati solo alla fine degli anni settanta, rendendo così possibile lo studio particolareggiato della loro azione.

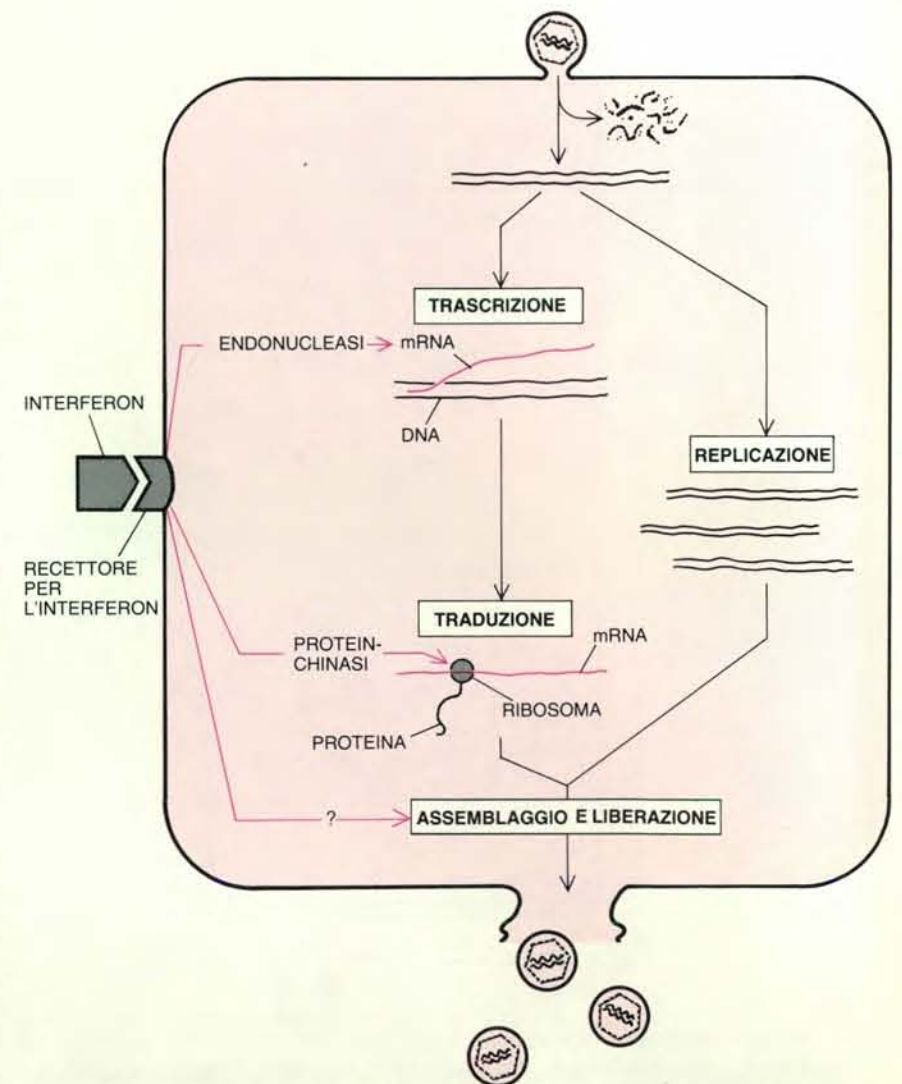
Le cellule producono interferon non solo quando si trovano direttamente esposte al virus, ma anche in presenza di RNA a duplice filamento (una forma di materiale genetico che è tipica di molti virus) o di certi composti che stimolano la risposta immunitaria. Le molecole di interferon così secrete si legano ai recettori di membrana di altre cellule, inducendo le cellule bersaglio a sintetizzare diverse nuove proteine.

Una di queste avvia una serie di reazioni biochimiche che porta all'attivazione di un enzima fino a quel momento quiescente, che inizia a demolire l'RNA messaggero del virus che si sta replicando. Vi è anche una seconda proteina, che è una chinasi, in grado di inattivare il meccanismo cellulare di sintesi proteica. Ciascuno di questi meccanismi (o entrambi, a seconda del tipo di cellula e di virus infettante) impedisce al virus di esprimere i propri geni in proteine. In alcune infezioni gli interferon producono anche altri effetti che non sono stati ancora spiegati a livello molecolare. Le microfotografie elettroniche mostrano, per esempio, un aumento di virus difettosi in alcune cellule stimolate da interferon, il che suggerisce che gli interferon possano inibire l'assemblaggio di nuove particelle virali e la loro liberazione.

Le possibilità terapeutiche offerte dagli interferon non hanno potuto essere messe alla prova fino a quando essi non sono stati disponibili in grandi quantità. Diversi anni dopo che le proteine erano state purificate per la prima volta, le tecniche del DNA ricombinante hanno permesso ai ricercatori di clonare i geni umani per l'interferon e di inserirli in batteri e lieviti. Questi organismi geneticamente modificati sono stati quindi capaci di produrre interferon umano in quantità virtualmente illimitate (si veda l'articolo *La purificazione e la produzione di interferon umani* di Sidney Pestka in «Le Scienze» n. 182, ottobre 1983).

Successivi studi clinici hanno suggerito che gli interferon potessero prevenire e curare alcune infezioni prodotte da virus erpetici e forse anche le verruche causate da papillomavirus. Rimane invece dubbia l'utilità degli interferon nell'epatite *B* cronica e nelle infezioni virali respiratorie. A causa dei risultati non conclusivi di molte prove cliniche e dei numerosi effetti tossici collaterali, gli interferon rimangono comunque in una sorta di limbo clinico.

La sfida dei nostri giorni: l'AIDS

Malgrado i progressi compiuti nel produrre farmaci antivirali o nel farli giungere al sito di infezione con effetti tossici collaterali minimi, la terapia antivirale

Gli interferon si legano a molecole che fungono da recettori sulla membrana superficiale di una cellula infetta (*a sinistra*) e possono quindi esercitare la loro azione antivirale in corrispondenza di diverse fasi del ciclo di replicazione del virus. Una di queste è la trascrizione del DNA virale, processo mediante il quale l'informazione genetica è copiata sotto forma di mRNA. Gli interferon possono avviare le reazioni biochimiche che attivano un'endonucleasi, enzima che demolisce l'RNA messaggero virale. Un altro possibile bersaglio è la traduzione, cioè la produzione di proteine dallo stampo di mRNA, che ha luogo sui ribosomi. Gli interferon possono stimolare la cellula a produrre una proteinchinasi che inattiva un fattore che prende parte alla traduzione. Un terzo bersaglio può essere la fase finale della replicazione virale: l'assemblaggio di nuove particelle virali a partire da proteine virali e da materiale genetico replicato e la loro liberazione. Il meccanismo con cui gli interferon bloccano l'assemblaggio e la liberazione del virus non è noto.

muove ancora i suoi primi passi. Alcune malattie virali, come la rabbia e le infezioni cerebrali provocate dagli arbovirus, si sono dimostrate finora incurabili. I progressi nella lotta contro altre infezioni, come il comune raffreddore e le diarree virali, sono stati scarsi. La massima sfida che la terapia antivirale deve affrontare rimane comunque quella di soddisfare l'impellente necessità di una cura efficace e duratura contro l'AIDS.

L'identificazione di questa sindrome risale ad appena sei anni fa; il suo agente etiologico, l'HIV, non è stato però identificato fino al 1983 (si veda l'articolo *Il virus dell'AIDS* di Robert C. Gallo in «Le Scienze» n. 223, marzo 1987). La malattia si è diffusa a una velocità impressionante e, nei soli Stati Uniti, vengono oggi considerate colpite da AIDS più di 30 000 persone. Il numero delle persone che, sempre negli Stati Uniti, hanno contratto l'infezione senza manifestarne ancora i sintomi è molto più elevato e verosimilmente supera il milione. Sono diverse le caratteristiche che consentono all'AIDS di resistere a ogni terapia. Il principale bersaglio del virus è il linfocito *T4*, un globulo bianco che mobilita le difese immunitarie. Pertanto, la distruzione dei linfociti *T4* nell'AIDS provoca una grave depressione della risposta immunitaria e l'eventuale farmaco si troverebbe a dover combattere il virus senza essere molto coadiuvato dall'immunità dell'ospite. Inoltre, il virus infetta anche le cellule del sistema nervoso centrale, dove rimane protetto dall'azione di molti farmaci che sarebbero potenzialmente efficaci se non ci fosse la barriera ematoencefalica.

Cionondimeno, sta prendendo corpo una serie di acquisizioni di base che, permettendo di capire il ciclo di replicazione del virus, appaiono decisive per la strategia antivirale nell'AIDS. Nell'infettare le cellule, l'HIV si lega a molecole specifiche, che sono i recettori di membrana. Il virus penetra così nel citoplasma e si libera dell'involucro proteico, esponendo il proprio materiale genetico, costituito da un singolo filamento di RNA. Un enzima virale, noto come transcriptasi inversa, è associato all'RNA. L'enzima in una fase che è caratteristica dei retrovirus (la classe di virus a cui appartiene l'HIV) «trascrive in senso inverso» l'RNA in DNA. Alla fine, alcune copie di DNA del genoma di HIV si integrano nei cromosomi della cellula ospite.

Il genoma virale integrato, noto come provirus, può rimanere latente fino a quando la cellula ospite viene stimolata, probabilmente da un'altra infezione. Il DNA del provirus viene allora trascritto in RNA messaggero che, a sua volta, controlla la sintesi delle proteine virali. Il provirus dà anche origine ad altre copie di RNA, che serviranno come materiale genetico della progenie virale. Le proteine e l'RNA genomico si raccolgono in prossimità della membrana cellulare e si uniscono per formare nuove particelle di HIV, che alla fine «gemmano» dalla cellula. Due geni del virus, indicati con le sigle *tat* e *trs/art*, governerebbero quest'esplosione di replicazione che finisce col distruggere la cellula. Essi codificano per piccole proteine che incrementano la trascrizione di DNA provirale e la sintesi di proteine virali.

Nel ciclo di replicazione dell'HIV, il bersaglio migliore per un farmaco è la trascrizione inversa, una fase che, se è indispensabile per la replicazione virale, non ha alcuna importanza per la cellula ospite. È già stato dimostrato che diversi farmaci riducono, *in vitro*, l'attività della transcriptasi inversa. La replicazione dell'HIV è molto più lenta in presenza di composti come la suramina (un agente antiparassitario), il tungstato di antimonio (detto anche HPA-23), il fosfonoformiato e una nuova classe di analoghi dei nucleosidi, chiamati bidesossinucleotidi. Tutti questi composti sono oggi utilizzati in prove cliniche su pazienti affetti da AIDS o da sindromi affini.

Un altro inibitore della transcriptasi inversa, l'azidotimidina (AZT), ha già dimostrato di essere clinicamente efficace. Come l'acyclovir, è un analogo di nucleoside e, in particolare, del nucleoside timidina; messo a punto negli anni sessanta come agente anticancro, l'AZT interferisce con la sintesi di nuovo DNA durante la trascrizione inversa, ponendo termine all'allungamento del filamento di DNA. Il composto, attivato da una triplice fosforilazione prodotta da chinasi cellulari, si lega alla transcriptasi inversa (per la quale ha un'affinità 100 volte superiore a quella che ha per le DNA-polimerasi cellulari). Alla fine, viene incorporato nella catena di DNA. In corrispondenza della posizione 3' dell'anello di zucchero, l'AZT possiede un gruppo azidico (N_3) al posto di un ossidrile. I nucleotidi che vengono aggiunti non trovano alcun punto di attacco e così la trascrizione inversa si blocca.

Gli effetti antivirali *in vitro* dell'AZT sono tanto selettivi quanto potenti: la concentrazione alla quale inibisce la replicazione virale è ben al di sotto del livello al quale, per esempio, comincia a danneggiare le funzioni immunitarie dei linfociti *T*. Alla luce degli incoraggianti risultati ottenuti in laboratorio, nel 1985-1986, un gruppo di studio del National Cancer Institute sotto la guida di Samuel Broder ha condotto le prime prove cliniche sull'AZT, studiando la farmacocinetica e la sicurezza del composto su un piccolo gruppo di pazienti affetti da AIDS o da quell'insieme di sintomi correlati all'AIDS noto come ARC. Successivamente, su 282 pazienti affetti da AIDS,sono state avviate indagini controllate con placebo e da protrarsi per sei mesi.

Alla sedicesima settimana dall'inizio dello studio, una commissione indipendente ha notato sorprendenti differenze nel tasso di mortalità tra i pazienti che avevano ricevuto un placebo e quelli ai quali veniva somministrato il farmaco; questi ultimi soffrivano anche di un minor numero di infezioni intercorrenti, erano aumentati di peso e le loro funzioni immunitarie erano migliorate. Di conseguenza, lo studio venne sospeso e a tutti i pazienti che avevano ricevuto il placebo fu somministrato AZT. È tuttavia necessaria una maggiore conoscenza del farmaco prima di poterlo prescrivere su vasta scala. Non è chiaro se i suoi effetti si dimostreranno duraturi o se il farmaco sarà altrettanto efficace nei primi stadi dell'infezione. La prospettiva di un uso a lungo termine ha anche accresciuto le preoccupazioni sulla considerevole tossicità dell'AZT per le cellule del midollo osseo, che sono i precursori delle cellule del sangue.

Anche altre fasi del ciclo di replicazione dell'HIV, diverse dalla trascrizione inversa, possono essere importanti bersagli di un attacco antivirale. L'interferon umano alfa che, come altri interferon, può inibire la produzione dell'RNA messaggero virale, la sintesi di proteine virali o l'assemblaggio e la liberazione di nuove particelle virali, rallenta effettivamente la replicazione *in vitro* dell'HIV ed è oggi sottoposto a prove cliniche. Anche il ribavirin, che agisce sulla sintesi proteica e su altri eventi del ciclo di replicazione virale, è stato utilizzato per prove cliniche in base alle quali sembra possa ritardare o impedire lo sviluppo dell'AIDS in individui sieropositivi, ma asintomatici.

Si possono anche cercare agenti antivirali che interferiscano con la fase iniziale dell'infezione - cioè l'attacco di una particella HIV a recettori di membrana specifici - oppure che colpiscano i prodotti proteici dei geni *tat* e *trs/art*, interferendo così con l'espressione di altri geni HIV. Questi agenti (per il momento in gran parte ipotetici) potrebbero essere altamente specifici e quindi meno tossici degli attuali farmaci antivirali.

Future strategie antivirali

La rapida comparsa sulla scena di promettenti terapie per l'AIDS è la dimostrazione di una più approfondita comprensione della replicazione virale a livello molecolare e il risultato di un'esperienza clinica ventennale con altri farmaci antivirali. Gran parte dei futuri progressi saranno sicuramente dovuti a una migliore utilizzazione dei farmaci esistenti. Batterie di esami per una diagnosi precoce delle infezioni virali permetteranno al medico di avviare il trattamento farmacologico prima che il danno cellulare si estenda. Metodi per inviare i farmaci a tessuti infettati oggi difficilmente raggiungibili, come per esempio il sistema nervoso centrale, porranno altre infezioni nel novero di quelle passibili di cura.

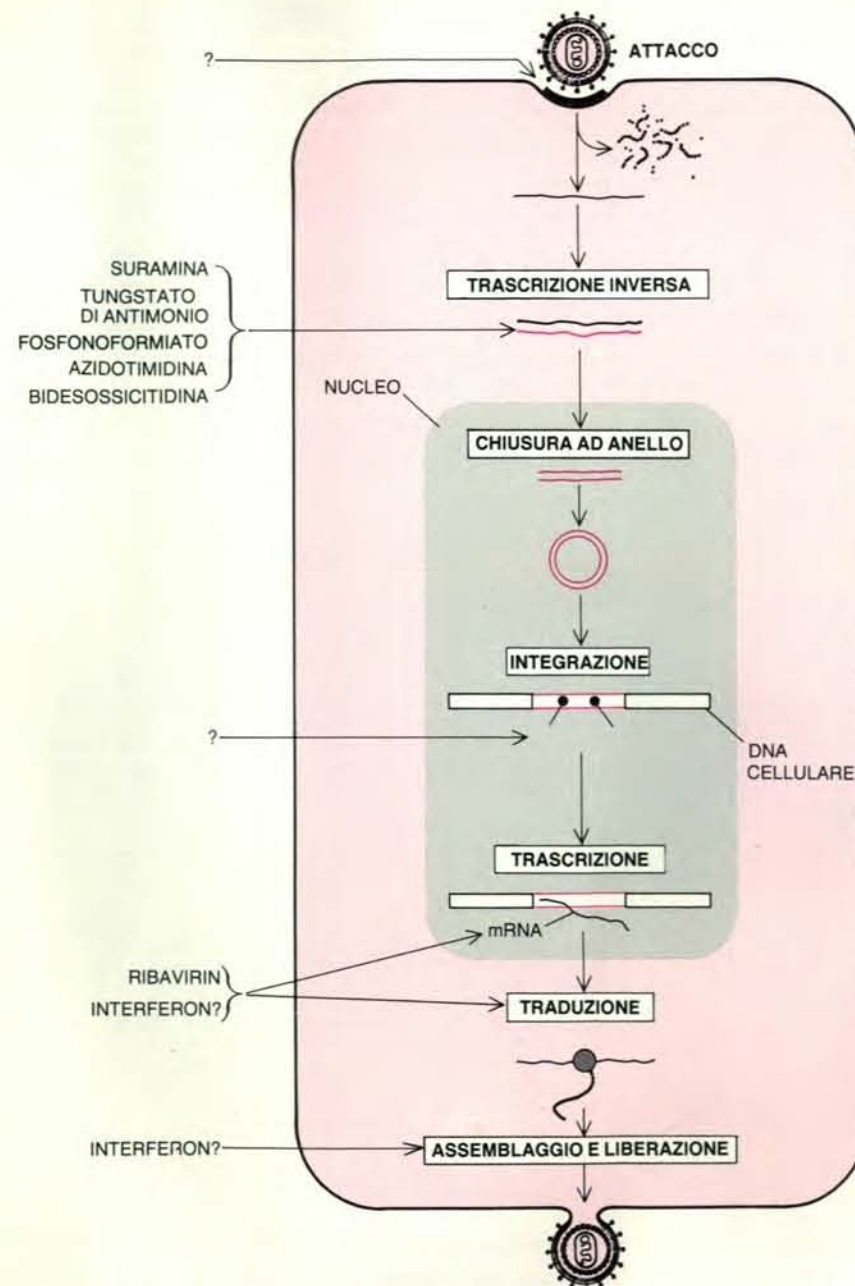
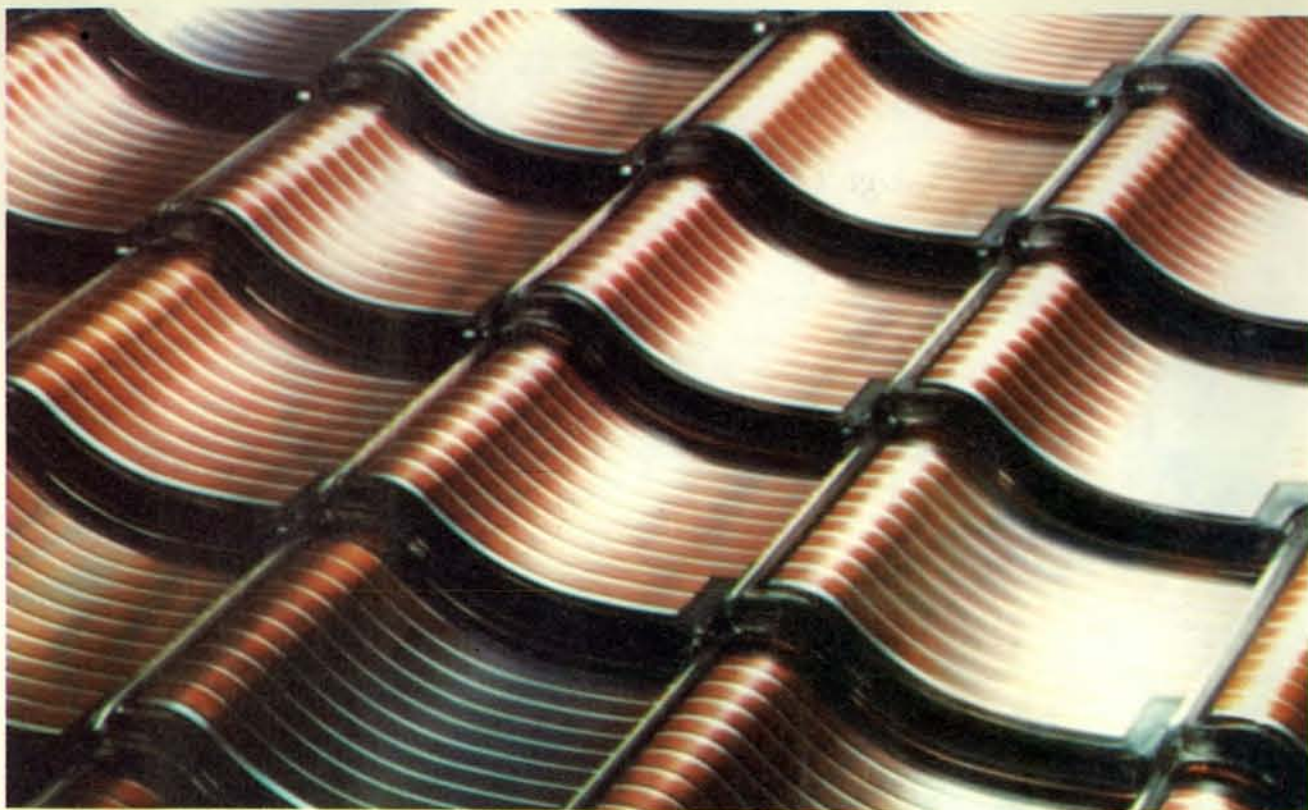
È verosimile che nell'AIDS e in altre gravi malattie virali a lungo decorso la terapia si appoggi sempre più ad associazioni di farmaci, ciascuno dei quali colpisca una differente funzione virale. Un possibile vantaggio di questa strategia sarebbe appunto il sinergismo antivirale. Per esempio, un gruppo di ricercatori ha dimostrato che la combinazione di un inibitore della transcriptasi inversa, come l'AZT o il fosfonoformiato, con l'interferon alfa ha, *in vitro*, un effetto sinergico sull'HIV .

Associazioni di questo tipo vengono oggi sperimentate su pazienti nei quali ci si potrebbero aspettare ulteriori benefici. Dato che ogni componente di queste associazioni è presente in quantità minore, è probabile che gli effetti tossici collaterali non compaiano. Inoltre, in questo modo viene ridotta di molto la probabilità che emerga un ceppo virale resistente: il virus dovrebbe infatti sviluppare mezzi per far fronte contemporaneamente a diversi farmaci. Vantaggi analoghi potrebbero anche venire da un protocollo terapeutico in cui due o più composti vengano somministrati in successione piuttosto che insieme. Nel caso dell'AIDS, una terapia basata sulle associazioni di farmaci antivirali potrebbe richiedere la contemporanea somministrazione di proteine naturali che regolano l'attività immunitaria, nel tentativo di ristabilire le difese del paziente.

La messa a punto di una associazione sinergica di agenti antivirali, al pari dell'identificazione di un composto promettente, esige una notevole attenzione ai meccanismi molecolari dei farmaci. Infatti, in associazione, alcuni farmaci potrebbero avere effetti antagonisti. Nel nostro laboratorio, per esempio, Markus Vogt ha dimostrato che il ribavirin ha un'affinità per gli enzimi cellulari che fosforilano l'AZT in forma attiva. L'AZT perde *in vitro* buona parte della sua efficacia quando è associato con il ribavirin. Pertanto, le prove cliniche su questa combinazione di farmaci dovrebbero essere affrontate con prudenza e realizzate in condizioni controllate.

Gli studi molecolari sui cicli di replicazione dei virus e sui meccanismi dei farmaci sperimentali rimarranno sicuramente la principale fonte di progresso terapeutico. Quando verranno identificate le vulnerabilità molecolari di altri virus, causa di infezioni finora incurabili, si riusciranno a produrre farmaci estremamente specifici. Studi molecolari sui virus latenti potranno forse dare origine anche a farmaci sufficientemente selettivi da colpire le sequenze di DNA virale silente, che si distinguono soltanto per la loro origine estranea al materiale genetico dell'ospite che le circonda.

Una strategia per lo sviluppo di nuovi farmaci, in futuro, potrà emergere anche dallo studio dei virus mediante cristallografia a raggi X, che può rivelarne la struttura nei particolari atomici (si veda l'articolo *La stuttura del poliovirus* di James M. Hogle, Marie Chow e David J. Filman in «Le Scienze» n. 225, maggio 1987). Sulla base di queste conoscenze strutturali, i programmi di grafica al calcolatore potrebbero fornire modelli per prevedere il sito di legame tra un composto sperimentale e una particella virale. Queste informazioni potrebbero quindi essere utilizzate per modificare il composto originale o per sintetizzare una nuova molecola che sia in grado di inserirsi in una determinata nicchia della struttura del virus. Il farmaco costruito su misura si legherebbe specificamente a un virus, impedendogli di attaccarsi alla cellula ospite o di liberarsi del proprio involucro proteico. Una simile strategia per la messa a punto razionale dei farmaci farebbe aumentare notevolmente le attuali possibilità della terapia antivirale.

ATTACCO
TRASCRIZIONE INVERSA
SURAMINA
TUNGSTATO DI ANTIMONIO
FOSFONOFORMIATO
AZIDOTIMIDINA
BIDESOSSICITIDINA
NUCLEO
CHIUSURA AD ANELLO
INTEGRAZIONE
DNA CELLULARE
TRASCRIZIONE
mRNA
RIBAVIRIN
INTERFERON?
TRADUZIONE
INTERFERON?
ASSEMBLAGGIO E LIBERAZIONE

I bersagli molecolari per farmaci anti-AIDS già esistenti e potenziali sono distribuiti lungo il ciclo di replicazione dell'HIV, l'agente etiologico dell'AIDS. Il virus infetta la cellula ospite legandosi a specifiche molecole della superficie cellulare; un farmaco antivirale efficace potrebbe agire bloccando l'attacco. All'interno della cellula, il virus si libera dell'involucro proteico e trasforma il proprio materiale genetico costituito da RNA in DNA attraverso un processo noto come trascrizione inversa e che è il bersaglio dei farmaci oggi in uso o in corso di studio. Il DNA virale entra quindi nel nucleo cellulare prendendo forma circolare; in parte viene poi incorporato, o integrato, nei cromosomi cellulari. Il DNA integrato è quindi trascritto in RNA messaggero, che viene poi tradotto in proteine virali. Geni noti con le sigle *tat* e *trs/art* specificano proteine che stimolano la trascrizione e la traduzione e potrebbero servire come bersagli per farmaci futuri. Un farmaco già disponibile che interferisce con la trascrizione e con la successiva maturazione dell'RNA messaggero è il ribavirin. Anche gli interferon possono inibire la trascrizione, e forse anche la traduzione e l'ultimo evento del ciclo: l'assemblaggio e la liberazione di nuove particelle virali.

Le tegole tradizionali possono essere sostituite con tegole fatte di celle fotovoltaiche di silicio amorfo, che danno la possibilità di sfruttare la luce del Sole per fornire all'edificio energia elettrica. Le tegole raffigurate sono state realizzate in Giappone dalla Sanyo Electric Co., Ltd.

Una casa alimentata con energia elettrica fotovoltaica è in funzione da cinque anni a Osaka. I pannelli azzurri sul tetto, fatti di celle di silicio amorfo, forniscono l'elettricità; i pannelli rossi inclinati posti intorno al tetto provvedono al riscaldamento e al raffreddamento dell'edificio.

Energia elettrica dal Sole

I rapidi progressi tecnologici realizzati di recente nella trasformazione di energia solare in energia elettrica fanno prevedere la costruzione, tra una quindicina d'anni, di centrali fotovoltaiche di vari megawatt di potenza

di Yoshihiro Hamakawa

È virtualmente certo che nei decenni a venire la domanda mondiale di energia manifesterà una crescita costante. Si può anche prevedere, sebbene con una certezza un po' minore a causa di certe variabili tecnologiche, che i produttori di energia dovranno affrontare sia la penuria di combustibili fossili, sia le preoccupazioni destate dagli effetti inquinanti di questi combustibili e dagli eventuali pericoli dell'energia nucleare. È evidente che una sorgente di energia basata su un combustibile inesauribile e non inquinante potrebbe assumere un ruolo preciso. Una sorgente siffatta è già disponibile oggi ed è costituita dalla cella fotovoltaica (o cella solare), che trasforma la luce del Sole in energia elettrica.

Questa tecnologia non è nuova; le celle fotovoltaiche hanno fornito elettricità ai veicoli spaziali fin dai loro primi voli e ora cominciano a essere utilizzate in certi prodotti commerciali, come le calcolatrici e gli orologi da polso. Il fatto nuovo è che si sta avvicinando il momento in cui con questa tecnologia si potranno costruire centrali elettriche da alcuni milioni di watt. Questi progressi sono dovuti tanto all'aumento del rendimento delle celle fotovoltaiche nella trasformazione della luce solare in energia elettrica, quanto alla diminuzione del loro costo di fabbricazione. In conseguenza di ciò, il costo dell'energia fotoelettrica è sceso nell'ultimo decennio da 50 a meno di cinque dollari per watt di picco. Sembra probabile che, grazie a ulteriori progressi, alla fine del secolo saranno in funzione grandi centrali elettriche basate su celle fotovoltaiche.

Il nome corretto del dispositivo che sta alla base di questa tecnologia è cella fotovoltaica. Fino a oggi virtualmente tutte le celle fotovoltaiche sono state costruite con silicio cristallino, ma esiste la possibilità di impiegare anche altri materiali; fra questi sono promettenti i semiconduttori fatti di certe leghe come il diseleniuro di rame e indio, il tellururo di cadmio e l'arseniuro di gallio.

Qualunque sia il materiale impiegato, il meccanismo con cui la cella trasforma la luce solare in energia elettrica è essenzialmente lo stesso. Il fenomeno che più interessa è il movimento dei portatori di carica nei semiconduttori, dato che l'elettricità, cioè la corrente elettrica, consiste appunto nel passaggio di queste cariche lungo un circuito. Questo processo può essere osservato, per esempio, in una cella fotovoltaica di silicio cristallino.

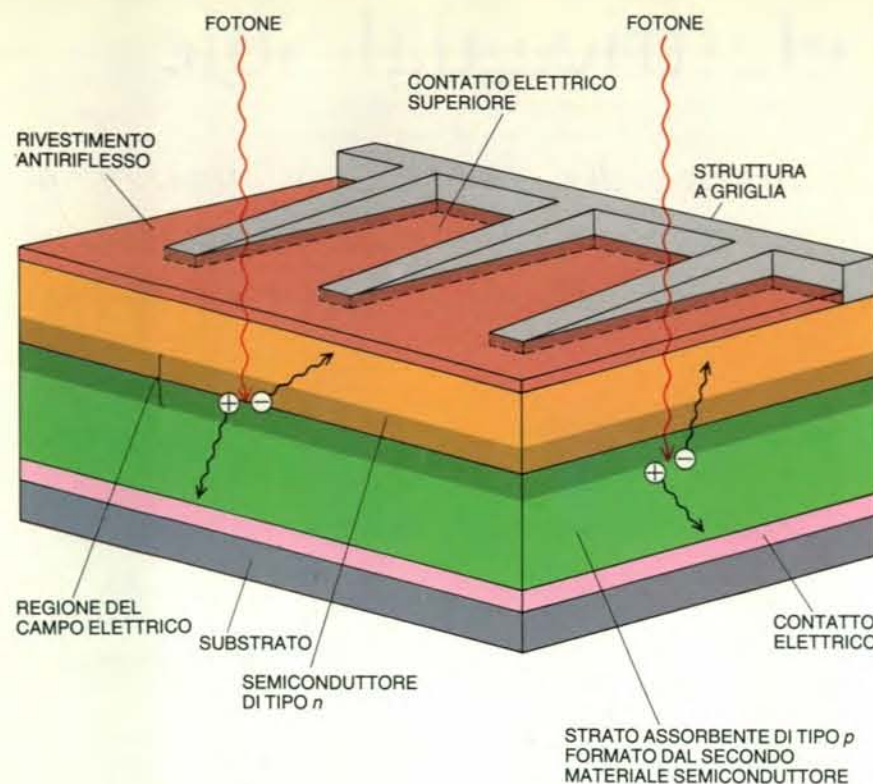
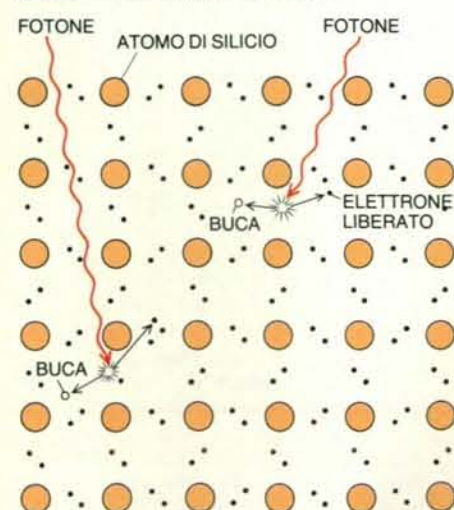
Normalmente l'atomo di silicio possiede 14 elettroni, quattro dei quali sono elettroni di valenza: ciò significa che essi possono partecipare alle interazioni con altri atomi, sia di silicio, sia di altri elementi. In un cristallo di silicio puro, ciascun atomo è legato in modo covalente con altri quattro atomi: una coppia di elettroni, uno dei quali appartenente all'atomo considerato e l'altro appartenente all'atomo vicino, viene messa in comune fra i due atomi che formano il legame.

Questo legame elettrostatico piuttosto forte fra un elettrone e i due atomi che esso contribuisce a tenere uniti può essere spezzato da una quantità sufficiente di energia. Infatti, se l'energia fornita è sufficiente, l'elettrone viene portato a un livello energetico superiore, chiamato banda di conduzione, dove è libero di spostarsi e può contribuire al flusso di elettricità. Quando passa alla banda di conduzione, l'elettrone si lascia dietro una «buca», cioè una lacuna dove manca un elettrone. Un elettrone vicino può andare facilmente a riempire la buca, scambiandosi così di posto con essa. Quindi all'interno di un cristallo, tanto gli elettroni quanto le buche si possono muovere.

Una cella fotovoltaica è sostanzialmente un diodo di grande superficie. In un diodo il passaggio della corrente, che è costituita da portatori di carica liberi, per esempio elettroni, è ostacolato in una direzione e facilitato in quella opposta. Questo risultato è ottenuto grazie alla presenza nel diodo di un campo elettrico fisso dovuto al fatto che il diodo è costituito da due semiconduttori di materiale diverso. Questo campo elettrico spinge le cariche di un segno ad attraversare il diodo e respinge le cariche di segno opposto.

Una cella fotovoltaica sfrutta proprio questo fenomeno: è costituita da strati di semiconduttori diversi che creano dentro la cella un campo elettrico fisso simile a quello del diodo. Questo campo «incorporato» è ubicato il più vicino possibile alla regione del dispositivo che assorbe la luce. La luce solare trasporta energia (circa 1000 watt per metro quadrato di superficie esposta). I fotoni, quando colpiscono un semiconduttore della cella fotovoltaica, possono venirne riflessi, attraversarlo, o esserne assorbiti. I fotoni potenzialmente utili sono quelli che vengono assorbiti.

Un fotone che venga assorbito da un semiconduttore può produrre calore urtando il materiale semiconduttore oppure, se possiede abbastanza energia, può strappare un elettrone da uno stato legato ed elevarlo a uno stato libero nella banda di conduzione del materiale. Per esempio, un fotone la cui lunghezza d'onda sia inferiore a circa 1,2 micrometri, qualora venga assorbito dal silicio, ha un'energia sufficiente per generare un elettrone libero.

Quando un elettrone passa alla banda di conduzione, si crea una buca: questa rappresenta una carica positiva libera che si comporta in modo analogo, ma opposto a una carica negativa libera. In un semiconduttore di silicio una buca può spostarsi liberamente perché un elettrone legato che si trovi nelle vici-

Una cella fotovoltaica che utilizza silicio cristallino è composta di più strati. Partendo dal basso s'incontra un substrato di vetro o di plastica; uno strato metallico che serve da contatto per il circuito elettrico esterno; un semiconduttore di silicio di tipo *p* che tende a raccogliere le «buche», cioè le particelle con carica positiva; uno strato di silicio di tipo *n*, che raccoglie invece gli elettroni, carichi negativamente; un rivestimento antiriflesso e una griglia, che costituisce il contatto superiore. I simboli più e meno rappresentano una coppia buca-elettrone creata dall'assorbimento di un fotone solare o quanto di luce. Vicino alla giunzione *p-n*, un campo elettrico prodotto dai diversi materiali dei semiconduttori spinge gli elettroni verso l'alto e le buche verso il basso. È questo flusso di cariche a generare la corrente elettrica nel circuito esterno.

L'effetto fotovoltaico dei fotoni in un cristallo di silicio (*a*) ha luogo quando un fotone incidente colpisce un elettrone (*puntini neri*) legato a un atomo di silicio. L'elettrone viene liberato e diviene una carica libera negativa, che lascia al suo posto una buca positiva. Questo è il fenomeno della fotoconduzione. Secondo la meccanica quantistica (*b*) gli elettroni liberati da fotoni di energia appropriata acquistano energia e passano dalla banda di valenza, cioè da uno stato legato, alla banda di conduzione. Se il fenomeno della fotoconduzione ha luogo vicino al campo elettrico della cella, la coppia buca-elettrone generata dalla luce viene separata e i suoi componenti vengono spinti verso i lati opposti della cella, dove possono contribuire alla corrente elettrica.

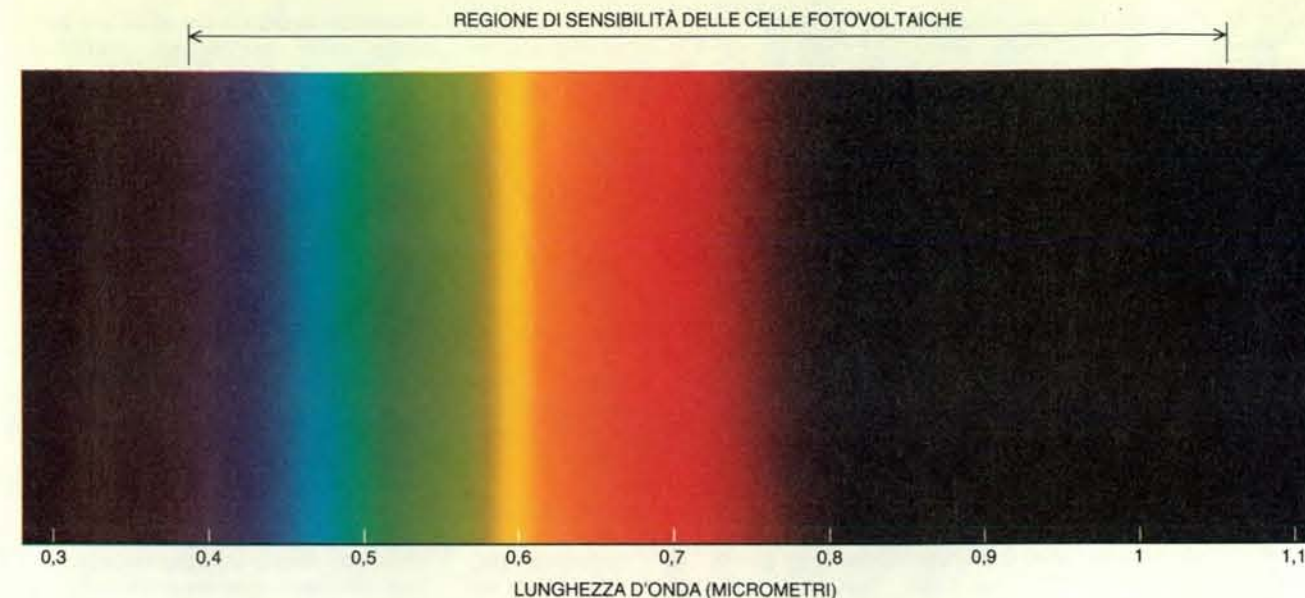
nanze può facilmente saltare nella posizione della buca, mentre quest'ultima va a occupare la posizione lasciata libera dall'elettrone che si è spostato.

La luce che viene assorbita da un semiconduttore produce due portatori di carica liberi: l'elettrone libero, nella banda di conduzione, e la buca libera, nella banda di valenza. La conversione da luce a enegia elettrica effettuata dalla cella fotovoltaica avviene essenzialmente perché questi portatori di carica liberi, generati dalla luce, sono spinti in direzioni opposte dal campo elettrico incorporato. Consideriamo l'esempio di buche ed elettroni che si trovano vicini al campo. Supponiamo che (rispetto a questi portatori generati dalla luce) il campo abbia un verso tale da facilitare il passaggio degli elettroni e respingere le buche. Gli elettroni allora attraversano il campo e vengono accelerati, mentre le buche restano vicino al punto dove sono state generate. Una volta che abbiano attraversato il campo, gli elettroni liberi non tornano indietro, perché il campo, agendo come un diodo, impedisce loro di invertire la marcia.

Perciò, quando la luce incide sulla cella fotovoltaica, le cariche positive sono spinte in numero crescente verso la parte superiore della cella e le cariche negative verso quella inferiore, o viceversa, a seconda del tipo di cella. Se la parte superiore e quella inferiore sono collegate da un conduttore, le cariche libere lo attraversano e si osserva una corrente elettrica. Fino a quando la cella resta esposta alla luce l'elettricità fluisce con regolarità sotto forma di corrente continua.

In genere una cella fotovoltaica poggia su un supporto di vetro o di plastica, che si chiama substrato e che serve a sostenere la cella. Su di esso è depositato uno strato conduttore, per esempio di metallo, che funge da polo inferiore. Poi viene depositato un semiconduttore fotoassorbente e, su quest'ultimo, viene posto un semiconduttore diverso o modificato. Il campo critico incorporato si trova in corrispondenza della superficie che separa questi due semiconduttori. Spesso il semiconduttore superiore è trasparente per consentire alla luce di passare e di essere assorbita il più vicino possibile al campo elettrico. Questo accorgimento accresce la probabilità che gli elettroni o le buche generati dalla luce raggiungano la zona del campo e vengano separati.

L'ultima componente di una cella è di solito una griglia che costituisce il polo superiore. Se i due poli sono tra loro collegati e formano un circuito, quando la cella viene illuminata si osserva il passaggio di una corrente elettrica che può essere sfruttata.

Per valutare le prestazioni di una cella fotovoltaica ben progettata è necessario tener conto di altri fenomeni importanti. Due processi, detti di deriva e di diffusione, influiscono sull'efficacia

La regione di risposta delle celle fotovoltaiche di silicio cristallino inizia nella zona violetta dello spettro visibile (a una lunghezza d'onda di 0,35 micrometri circa) e lo comprende tutto, giungendo fino all'infrarosso. L'energia solare ha intensità massima tra 0,5 e 0,55 micrometri.

con cui i portatori liberi generati dalla luce contribuiscono alla corrente. La luce assorbita dalla cella può generare portatori liberi sia nella stretta regione del campo, sia fuori di essa. I portatori generati dentro questa regione, che normalmente ha una larghezza di 0,5 micrometri, vengono separati molto efficacemente dal forte campo elettrico; poi sono spinti ad alta velocità verso la parte superiore e verso la parte inferiore della cella. Questo è il processo di deriva.

La diffusione interviene sui portatori liberi generati fuori della regione del campo. Essa dipende dalla loro tendenza a rimbalzare qua e là, casualmente, per un certo tempo. Un elettrone libero ha una certa probabilità di finire nel campo elettrico incorporato: quando questo accade, l'elettrone viene spinto dall'altra parte della cella e può contribuire alla corrente elettrica. Le celle solari di elevata qualità sono progettate per rendere massima questa probabilità.

Lo spessore di una cella fotovoltaica e, in larga misura, il suo costo in termini di materiali e di lavorazione sono determinati dalla sua capacità di assorbire la luce solare. Questo assorbimento dipende da due fattori critici, l'ampiezza della banda proibita e il coefficiente di assorbimento. La banda proibita di un semiconduttore è il livello di energia caratteristico al quale il semiconduttore inizia ad assorbire la luce. Materiali come il silicio, l'arseniuro di gallio e altri vengono impiegati nelle celle fotovoltaiche perché normalmente iniziano ad assorbire la luce già a elevate lunghezze d'onda: di conseguenza riescono ad assorbire una quota cospicua della luce ambiente.

Inoltre i diversi semiconduttori presentano un assorbimento ottimale in corrispondenza di spessori diversi, che possono andare da 100 micrometri a meno di un micrometro. Le cosiddette celle fotovoltaiche a pellicola sottile sono fatte di materiali che presentano il massimo assorbimento a uno spessore intorno a un micrometro, il che consente di ridurre molto i costi di materiali e lavorazione. (Occorrono circa 50 micrometri di silicio cristallino per assorbire la quantità di luce che può essere assorbita da un micrometro di silicio amorfo e da 0,2 micrometri di diseleniuro di rame e indio.)

Queste differenze di assorbimento sono dovute alle differenze di struttura atomica dei diversi semiconduttori. Alcuni materiali, detti a banda proibita diretta, presentano un forte assorbimento; altri, detti a banda proibita indiretta, assorbono poco. Il silicio cristallino è un materiale a banda proibita indiretta; i materiali per le pellicole sottili, come il silicio amorfo e il diseleniuro di rame e indio sono materiali a banda proibita diretta. Nei materiali di tipo indiretto, affinché un elettrone di valenza passi alla banda di conduzione è necessario il verificarsi simultaneo di due fenomeni: l'assorbimento luminoso e l'interazione favorevole con una vibrazione reticolare (detta fonone). Nei materiali a banda proibita diretta, invece, è sufficiente l'energia della luce. La differenza fra la probabilità di quest'unico evento e la probabilità dei due eventi sopra menzionati spiega perché fra i due tipi di materiali l'assorbimento luminoso differisca di un ordine di grandezza.

Quando si lavora con le celle fotovoltaiche, si deve anche tener conto della quantità di luce solare che raggiunge la Terra. Appena al di sopra dell'atmosfera terrestre, la costante solare è pari a 1358 watt per metro quadrato. Quando la luce solare raggiunge il livello del mare, l'assorbimento da parte del vapore acqueo, dell'ozono (O_3) e dell'anidride carbonica ne riduce la densità di potenza a circa 1000 watt per metro quadrato. L'energia solare, a livello della superficie terrestre, è concentrata nell'intervallo di lunghezza d'onda da 0,4 a 2,5 micrometri; l'energia ha un massimo fra 0,5 e 0,55 micrometri, la lunghezza d'onda del colore delle foglie.

Il rendimento di una cella è determinato dallo spettro di energia dell'assorbimento ottico dell'area attiva del semiconduttore e dalla profondità della giunzione *p-n* (la superficie che separa lo strato che tende a raccogliere le cariche positive, cioè le buche, e lo strato che tende a raccogliere le cariche negative, cioè gli elettroni) misurata a partire dalla faccia della cella esposta alla luce. Un esempio del grado di efficienza che ci si può aspettare da una cella fotovoltaica è fornito dalla cella di silicio cristallino. Queste celle sono sensibili alla luce nell'intervallo fra circa 0,35 e circa 1,1 micrometri, che contiene quasi tutto lo spettro visibile e si prolunga nell'infrarosso. A causa delle perdite dovute a fattori quali il calore, la riflessione e i fotoni che non vengono assorbiti dal semiconduttore, il massimo rendimento teorico della conversione fotoelettrica di una cella di silicio è circa il 28 per cento. Altre perdite all'interno della cella fanno sì che il rendimento effettivo sia leggermente inferiore.

Viene spesso sottolineato che il rendimento di conversione di una cella fotovoltaica è meno della metà di quello dei sistemi di generazione basati sull'energia nucleare e sui combustibili fossili. Questo confronto è privo di senso. Infatti una turbina a vapore alimentata a petrolio che abbia un rendimento del 38 per cento sciupa il 62 per cento di un combustibile costoso. (Anzi il combustibile

deriva da processi dipendenti dall'energia solare molto meno efficienti di una cella fotovoltaica.) Per di più l'energia generata provoca inquinamento termico, piogge acide e sporca l'atmosfera. D'altro canto, se un'installazione di celle fotovoltaiche avesse un rendimento di conversione fotoelettrica anche solo del 10 per cento, sfrutterebbe una sorgente di energia che non costa nulla e che altrimenti non verrebbe utilizzata.

Ciò non significa che non si debba cercare di migliorare il rendimento delle celle fotovoltaiche. Ogni aumento di rendimento contribuisce a ridurre il costo della potenza di uscita degli impianti fotovoltaici. Quindi molte ricerche vengono condotte allo scopo di migliorare il rendimento di conversione (da luce a elettricità) delle tecnologie candidate, in modo da ottenere potenze superiori.

Un accorgimento ovvio ma importante consiste nell'applicare alla superficie anteriore della cella un rivestimento antiriflettente, in modo da aumentare la percentuale di luce non riflessa e quindi il numero dei fotoni che penetrano nella giunzione *p-n*. In una cella non rivestita le perdite per riflessione possono raggiungere il 30 per cento. Si cerca anche di impiegare la cosiddetta tecnologia del campo superficiale posteriore, che consiste nel creare un campo elettrico presso la superficie posteriore di un semiconduttore in modo da respingere verso la giunzione *p-n* i portatori di carica liberi in movimento casuale e migliorare così la separazione di elettroni e buche. Grazie a questi accorgimenti, il rendimento delle celle fotovoltaiche cristalline di serie ha raggiunto circa il 14 per cento.

Di recente sono stati sperimentati parecchi nuovi materiali per celle fotovoltaiche. Uno dei più importanti è il silicio amorfo (a-Si), che non possiede la struttura reticolare regolare del silicio cristallino. La struttura amorfa accresce notevolmente la probabilità che la luce venga assorbita e che gli elettroni passino nella banda di conduzione. Dal punto di vista dell'assorbimento luminoso, questo materiale presenta dunque notevoli vantaggi rispetto al silicio cristallino. Se ne possono quindi ricavare semiconduttori a pellicola sottile (con uno spessore di 0,5 micrometri rispetto ai 300 micrometri del silicio cristallino). Dato che contengono una minore quantità di materiale, le celle a pellicola sottile sono più economiche.

A questa prospettiva si collega l'introduzione del carburo di silicio amorfo (a-SiC). In questa lega il costruttore può modificare non solo il numero di elettroni di valenza (drogando il materiale con impurezze capaci di aumentare o diminuire il numero degli elettroni), ma anche l'ampiezza della banda proibita - cioè la quantità di energia incidente necessaria per portare un elettrone nella banda di conduzione - alla quale la cella reagisce (variando la quantità di carbonio). Questa scoperta inaugura l'era delle leghe amorfe di silicio, caratterizzata da materiali nuovi quali il silicio-germanio amorfo (a-SiGe), il nitruro di silicio amorfo (a-SiN) e altri ancora. Queste leghe modificano la banda proibita del silicio amorfo e la rendono quindi adattabile alle diverse funzioni previste dal progetto delle celle a pellicola sottile.

Un altro progresso compiuto nella tecnologia delle celle fotovoltaiche a pellicola sottile consiste nel sovrapporre più celle, per sfruttare meglio lo spettro solare. Per esempio, la cella superiore potrebbe raccogliere la luce di frequenza più elevata e lasciare passare quella di frequenza più bassa a un'altra cella, più efficiente in quella porzione dello spettro.

Un pannello sigillato contenente un certo numero di celle collegate tra loro già in sede di fabbricazione si chiama modulo fotovoltaico. Più moduli costituiscono una schiera. Per raccogliere la luce solare con maggiore efficienza, si può fare in modo che la schiera sia disposta su un piano; questa configurazione, detta piatto, è a sua volta montata su una struttura di sostegno che può essere progettata in modo da seguire il movimento del Sole durante le ore diurne. Un altro accorgimento per raccogliere la luce del Sole è adottato nella configurazione detta a concentratore, in cui si usano riflettori o lenti per concentrare la luce sulle celle fotovoltaiche. La differenza fra i due tipi di configurazione è che quella a piatto funziona in modo soddisfacente anche con la luce diffusa di una giornata coperta, mentre quella a concentratore funziona bene solo se illuminata dalla luce diretta del Sole. Questo svantaggio della configurazione a concentratore può essere in parte compensato impiegando nel sistema celle fotovoltaiche ad alto rendimento, per esempio celle a cristallo unico di arseniuro di gallio e silicio.

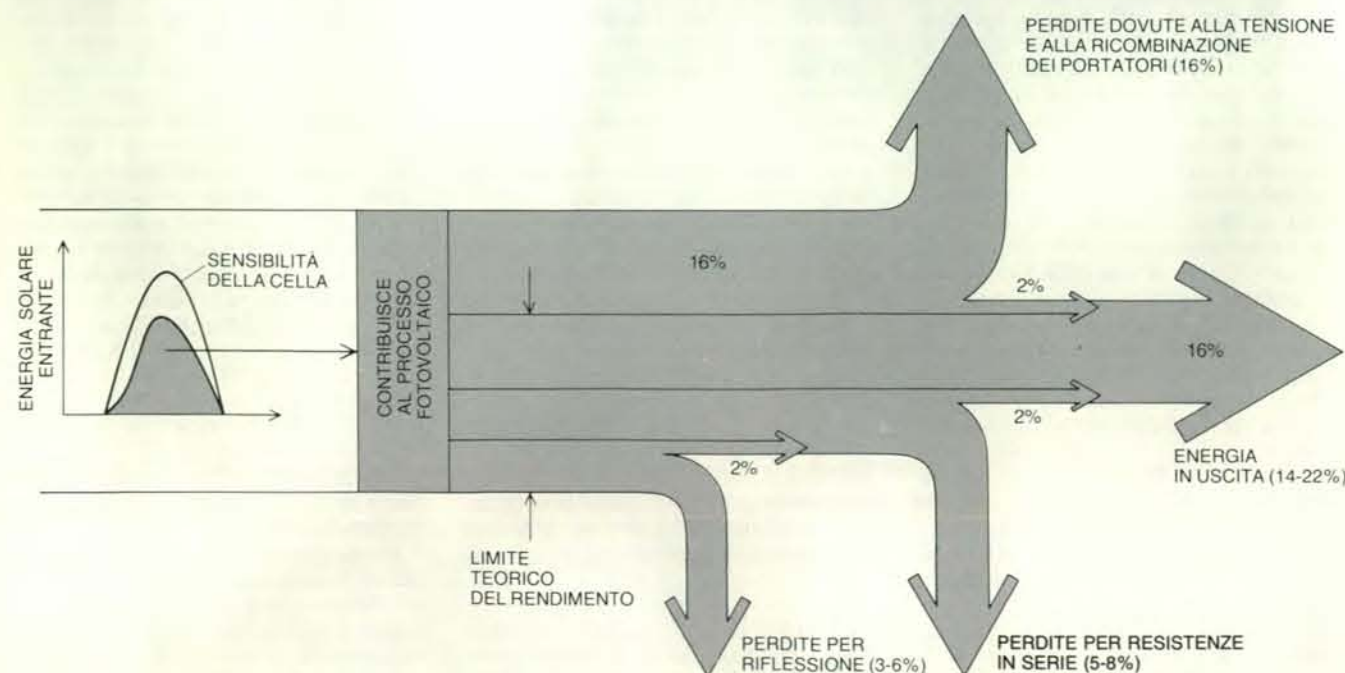
L'impiego di una configurazione a concentratore può contribuire a ridurre il costo degli impianti fotovoltaici. I dispositivi ottici come le lenti e i riflettori sono spesso più economici delle celle fotovoltaiche e, quindi, una centrale che li utilizzi richiede minori investimenti in dispositivi fotovoltaici. Inoltre in un impianto a concentratore la cella fotovoltaica di piccola superficie situata nel fuoco della concentrazione non è l'elemento decisivo per il costo del sistema; perciò è possibile utilizzare dispositivi fotovoltaici molto efficienti, il cui costo sarebbe eccessivo per una configurazione a piatto.

Il costo è il fattore che ha ritardato l'avvento degli impianti fotovoltaici di grandi dimensioni. Nel 1974 il costo di un modulo fotovoltaico era di 50 dollari per watt di potenza di picco. Il costo dell'elettricità ricavata da un modulo del genere sarebbe stato, sulla rete, elevatissimo: tre dollari per chilowattora, supponendo che la durata dell'impianto fosse di 20 anni e aggiungendo altri 50 dollari per watt di picco per tener conto del costo degli altri componenti del sistema. A quel tempo il costo dell'energia elettrica ottenuta dai combustibili fossili o nucleari variava dai cinque ai dieci centesimi di dollaro per chilowattora. Questa cospicua differenza imponeva tremendi sforzi di ricerca per diminuire il costo dell'energia elettrica fotovoltaica. Tali sforzi si sono concentrati sui materiali delle celle, la loro struttura e i procedimenti di produzione in serie e hanno avuto il risultato di far diminuire il costo di un modulo di un ordine di grandezza, fino a circa cinque dollari per watt di picco. Nel prossimo decennio ci si può aspettare un'ulteriore riduzione, forse di un altro ordine di grandezza: il costo potrebbe scendere, per le celle a pellicola sottile, per esempio di silicio amorfo, al di sotto dei 40 centesimi di dollaro per watt di picco.

Per illustrare la potenzialità delle celle solari, consideriamo la tecnologia delle pellicole sottili. Celle realizzate con tre dei materiali usati per costruire le pellicole sottili (silicio amorfo, diseleniuro di rame e indio e tellururo di cadmio) hanno raggiunto, in laboratorio, un rendimento del 10 per cento. Negli Stati Uniti e in Giappone parecchie società stanno costruendo moduli di silicio amorfo per il mercato, la cui superficie varia in certi casi da 1000 a 4000 centimetri quadrati. Il rendimento di questi moduli è circa dell'8 o 9 per cento. Alcune società stanno mettendo in commercio gli altri due materiali, che presentano le stesse potenzialità del silicio amorfo per quanto riguarda il rendimento a lungo termine e i livelli di costo richiesti dalla tecnologia delle pellicole sottili. Usando il diseleniuro di rame e indio, è stato possibile ottenere rendimenti quasi del 10 per cento, in moduli aventi area di 100 centimentri quadrati. Parecchie società impegnate con il tellururo di cadmio hanno sperimentato il metodo dell'elettrodeposizione per costruire celle fotovoltaiche di basso costo. Sono già stati costruiti moduli di tellururo di cadmio con una superficie di un migliaio di centimetri quadrati o più.

Le ricerche intese a migliorare questi materiali, in modo che sia possibile produrre energia a prezzi concorrenziali, hanno alcuni obiettivi precisi: aumentare il rendimento dei moduli fino a circa il 15 per cento; sviluppare tecniche di costruzione che consentano di ottenere moduli di costo inferiore ai 50 dollari per metro quadrato; garantire che i moduli abbiano lunga durata e funzionino bene negli impianti all'aperto per una trentina d'anni. Il fatto che i dispositivi a pellicola sottile abbiano conseguito un rendimento di conversione del 10 per cento (partendo da zero solo dieci anni fa) indica la rapidità dei progressi. Parecchie società affermano di aver raggiunto in laboratorio rendimenti dell'11 o 12 per cento. A causa delle perdite intrinseche, se si vuole che il rendimento medio delle schiere di grandi dimensioni sia del 15 per cento, il rendimento delle celle installate dovrà essere, in ultima analisi, del 16 o 17 per cento.

I progressi verso questi livelli di rendimento sono continui e vengono sistematicamente sviluppati nuovi metodi per realizzarli (per esempio la sovrapposizione delle celle). La possibilità di costruire celle di grande superficie sembra ormai a portata di mano. Quando i nuovi materiali entreranno in produzione su vasta scala assisteremo a cospicue riduzioni dei costi.

Via via che, negli ultimi dieci anni, il costo delle celle fotovoltaiche scendeva, sono comparse parecchie nuove applicazioni di questa tecnologia. Oltre agli svariati dispositivi elettronici funzionanti a celle fotovoltaiche oggi in vendita, ricordiamo le pompe idrauliche e i dispositivi di segnalazione per il traffico automobilistico e ferroviario. Generatori leggeri alimentati da celle fotovoltaiche forniscono elettricità a imbarcazioni e a veicoli terrestri, come i carrelli da golf. Come ha osservato Kenneth Zweibel, del Solar Energy Research Institute,

Il flusso di energia in una cella fotovoltaica di silicio cristallino sfrutta in linea di principio circa il 44 per cento dell'energia solare cui la cella è sensibile. Il 16 per cento circa dell'energia va perduta a causa di processi interni alla cella e ciò abbassa il rendimento a un limite teorico del 28 per cento. Altre perdite riducono il rendimento effettivo della conversione fotovoltaica, portandolo fra il 14 e il 22 per cento. Le frecce contrassegnate «2 per cento» indicano energia perduta, per esempio per riflessione, che può essere recuperata se la cella è progettata bene.

L'illustrazione evidenzia la differenza strutturale fra il silicio cristallino (*a*) e il silicio amorfo (*b*). Fino a tempi recenti le celle solari erano realizzate soprattutto con il silicio cristallino, che possiede la struttura reticolare regolare tipica dei cristalli. La struttura amorfa, irregolare, accresce notevolmente la probabilità che la luce sia assorbita, poiché i fotoni interagiscono più attivamente con la struttura amorfa. Quindi si può costruire una cella con meno materiale, ricorrendo alla cosiddetta tecnologia a pellicola sottile. Il silicio amorfo è uno dei numerosi materiali sperimentati di recente per le celle a pellicola sottile; fra gli altri materiali vanno ricordati l'arseniuro di gallio e il tellururo di cadmio.

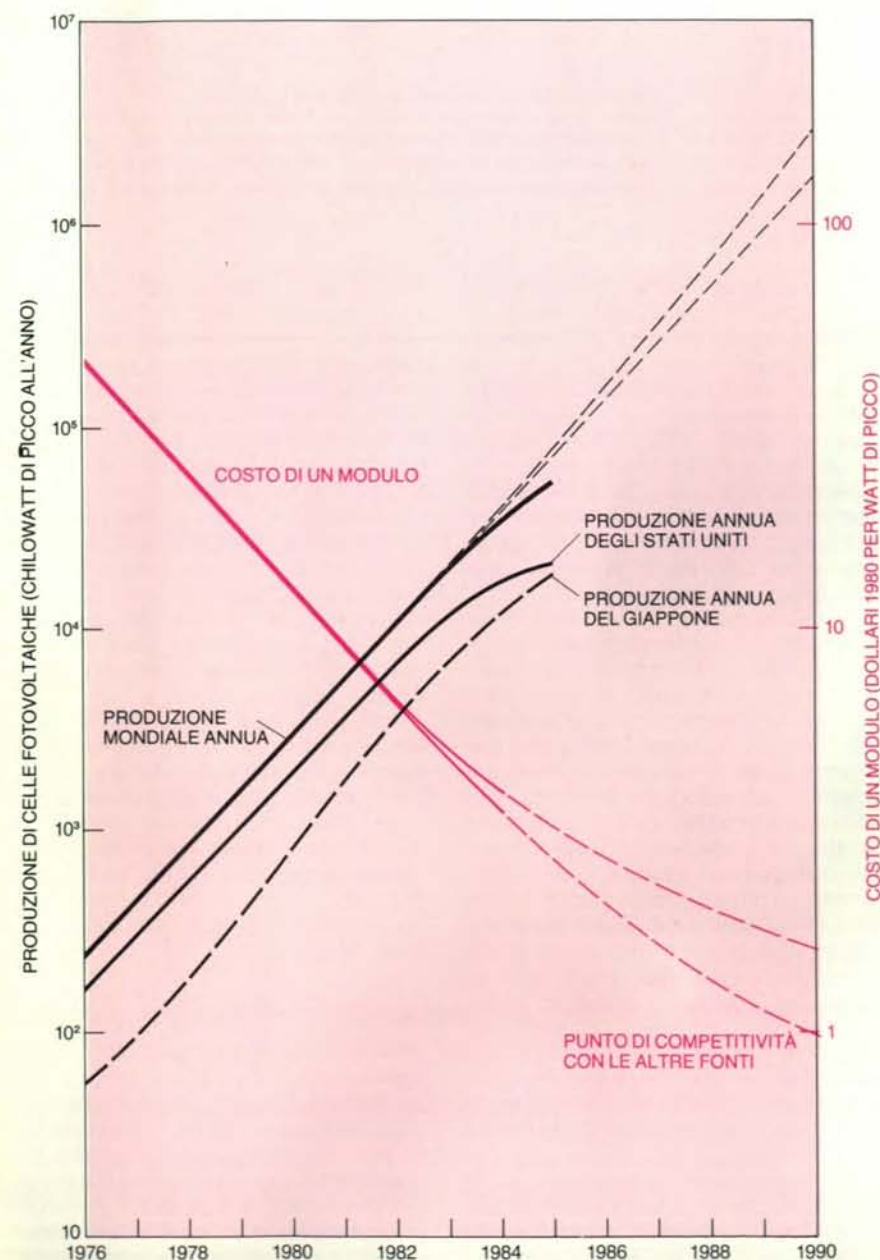
questa tecnologia dovrebbe trovare numerose applicazioni nei paesi in via di sviluppo.

L'obiettivo successivo è la costruzione di un sistema fotovoltaico di grandi dimensioni allacciato alla rete di distribuzione. Per consentire prove e dimostrazioni in tale prospettiva, verso la fine degli anni settanta il Department of Energy degli Stati Uniti erogò un finanziamento per nuove centrali, la cui potenza andava da 10 chilowatt a 100 megawatt. Questi impianti hanno dimostrato che è possibile costruire centrali di dimensioni intermedie e anche maggiori. Nel 1984, gli impianti fotovoltaici, esistenti o progettati, di dimensioni abbastanza grandi erano 14. La loro potenza va da 200 chilowatt a 120 megawatt e si trovano, fra l'altro, negli Stati Uniti, in Italia, in Giappone, in Arabia Saudita e nella Germania Federale. La Comunità Europea e alcune singole nazioni stanno costruendo 20 impianti fotovoltaici sperimentali da 15 a 300 chilowatt di potenza. In Giappone funzionano più di 10 impianti da tre a 200 chilowatt e un impianto da un megawatt, realizzato dalla New Energy Development Organization, è entrato in funzione nel 1985. Il progetto giapponese Sunshine, iniziato nel 1974, mira a sviluppare le fonti energetiche rinnovabili allo scopo di soddisfare una considerevole porzione della domanda energetica del paese per l'anno 2000. Da parecchi anni il Department of Energy degli Stati Uniti persegue un programma avente gli stessi scopi.

Il più grande impianto fotovoltaico oggi in funzione si trova a Carissa Plains, in California. È stato costruito in meno di un anno dell'ARCO Solar, Inc. e dalla Pacific Gas & Electric Company e fornisce 7,2 megawatt di potenza di picco alla rete di distribuzione.

Tenendo conto dei miglioramenti previsti nei materiali e nelle prestazioni, l'energia elettrica ottenuta dagli impianti fotovoltaici potrebbe avere un costo competitivo rispetto a quello delle centrali tradizionali già nel 1995. Il costo per chilowattora dell'elettricità prodotta dagli impianti tradizionali varierà dai tre centesimi (centrali idroelettriche) ai 35 centesimi di dollaro (turbine a vapore alimentate a carbone), mentre gli impianti fotovoltaici dovrebbero comportare un costo attorno agli otto centesimi al chilowattora. I successivi perfezionamenti della tecnologia dovrebbero permettere di vendere energia a meno di quattro centesimi al chilowattora.

Oltre alla pulizia e all'inesauribilità della sua sorgente, un sistema fotovoltaico presenta numerosi altri vantaggi. Non ha parti in movimento e quindi una centrale solare non ha virtualmente bisogno di manutenzione: ciò si è dimostrato vero nei fari comandati a distanza, nelle centrali di commutazione telefonica e nei veicoli spaziali. Le celle fotovoltaiche possono facilmente essere prodotte in serie come i transistori e i circuiti integrati a semiconduttore. In un mercato in espansione ci si può aspettare che questa caratteristica possa comportare una riduzione dei costi per effetto delle economie di scala. Le schiere solari sono fatte di numerosi moduli di celle e consentono una vasta gamma di dimensioni e di tipi per le varie applicazioni, con lo stesso rendimento di conversione e la stessa tecnologia. Un impianto da un megawatt può essere costruito in meno di un anno, e questa è una grande differenza rispetto alle centrali tradizionali e in particolare a quelle nucleari. Di conseguenza la pianificazione delle compagnie produttrici di elettricità potrà adeguarsi con più flessibilità alle proiezioni di variazione della domanda di energia elettrica.

Alla luce di tutte queste considerazioni, ci si può aspettare che l'era dello sfruttamento economico dell'energia fotovoltaica si aprirà nel prossimo futuro, molto prima di quanto non si sarebbe potuto supporre dieci anni fa. Questo traguardo richiederà un intenso lavoro di ricerca e di sviluppo, una collaborazione internazionale e forse qualche sussidio iniziale da parte dei singoli governi. In condizioni favorevoli, l'industria fotovoltaica potrebbe con il tempo raggiungere dimensioni tali da competere con quella elettronica.

La produzione e il costo delle celle fotovoltaiche hanno avuto nell'ultimo decennio andamenti opposti e si prevede che la tendenza continuerà. Un modulo è un pannello sigillato contenente celle fotovoltaiche collegate tra loro in sede di fabbricazione; più moduli compongono una schiera.

L'acustica delle antiche campane cinesi

Un carillon di campane di bronzo rinvenuto dagli archeologi ha rivelato la raffinata progettazione acustica di questo strumento orchestrale, che ebbe grande importanza prima di svanire dalla storia circa 2000 anni fa

di Sinyan Shen

Nel 1978, nella provincia dello Hubei, nella Cina meridionale, è stato recuperato un insieme di campane cinesi da carillon in bronzo, tanto grande da occupare tutto il palcoscenico di una sala da concerto moderna. Il carillon, che risale al V secolo a.C., è costituito da 65 campane che coprono cinque ottave, un intervallo superiore a quello della maggior parte degli strumenti moderni. Una filigrana di iscrizioni in oro sulle campane e il loro telaio documentano l'esistenza di una complessa teoria musicale che specificava il progetto, le dimensioni e la strumentazione delle antiche orchestre. Questa documentazione e ricerche recenti sulle campane stesse ci costringono a riscrivere daccapo la storia dell'acustica.

Le iscrizioni antiche hanno confermato quello che gli studiosi cominciavano appena a sospettare, ossia che le campane fossero costruite in modo tale che ciascuna potesse produrre due note distinte. Questa caratteristica distingue nettamente le campane cinesi da carillon dalle campane da chiesa dell'Occidente, che sono conosciute per la loro unica nota di grande durata; a differenza delle campane da chiesa, le campane da carillon possono eseguire musica dal ritmo complesso e veloce. Per motivi che non ci sono chiari, i principi e le pratiche relative a questo tipo di campane, rimasto unico nella storia, non sono mai stati tramandati. Di conseguenza la tecnica esecutiva delle campane da carillon è rimasta un mistero per più di 2000 anni.

Le campane incorporano molte caratteristiche fuori dal comune, la cui raffinatezza e la cui precisione possono essere valutate appieno soltanto facendo riferimento al progetto a due note. Dopo la scoperta del carillon, gli studi sulle proprietà vibrazionali e sui metodi di accordatura hanno messo in luce la profondità delle conoscenze possedute dai metallurgi e dai musicisti cinesi. Il progetto delle campane richiede una conoscenza teorica della fisica e dell'ingegneria che in precedenza si pensava si fosse sviluppata solo verso la fine del XVIII secolo. Invece i principi di acustica sfruttati nelle campane di bronzo cinesi hanno stupito persino gli studiosi di acustica del XX secolo.

Lo studio dell'acustica nel mondo occidentale ha origini relativamente recenti. Nel 1787 il fisico tedesco Ernst F. Chladni sparse della sabbia su lamine vibranti per evidenziare come certe regioni delle lamine rimanessero stazionarie durante la vibrazione. Queste regioni stazionarie sono state chiamate linee nodali e la loro distribuzione descrive i modi che assume un corpo in vibrazione. Ogni modo cosiddetto normale è associato a una frequenza di vibrazione caratteristica, e la frequenza di vibrazione determina la nota, cioè l'altezza del suono percepito.

I corpi in vibrazione si muovono in molti modi diversi simultaneamente e generano molte componenti diverse di frequenza, chiamate parziali. La parziale di frequenza più bassa è detta fondamentale; esistono molte frequenze superiori, che sono chiamate sovratoni. Quando si percuote una campana o si pizzica una corda, entrano in gioco tutte queste frequenze, ma alcune hanno intensità maggiore di altre. L'intensità relativa delle parziali in un suono musicale costituisce la sua qualità timbrica, esattamente come una combinazione di lunghezze d'onda determina il colore della luce.

Nel 1890 Lord Rayleigh studiò le campane nella torre campanaria della chiesa di Terling, in Inghilterra, e svolse alcuni esperimenti su numerose campane nel suo laboratorio. Identificò nelle campane sei parziali. Rayleigh, che pose i fondamenti delle ricerche successive sull'acustica delle campane, pensava che una campana potesse produrre solamente una nota fondamentale. Poiché la sua esperienza si limitava alle campane occidentali, non avrebbe potuto prevedere la lezione che ci avrebbero impartito campane cinesi di migliaia di anni fa.

Circa 80 anni fa, fra gli altri reperti archeologici, in Cina cominciarono ad apparire singole campane da carillon. Poi se ne scoprirono gruppi e interi carillon, al punto che oggi sono state recuperate migliaia di campane e oltre 50 carillon. Queste campane sono state esaminate attentamente, ma i ricercatori non si sono resi conto della loro caratteristica di poter produrre due note distinte fino al 1977. Gli ultimi dubbi sono stati spazzati via nel 1978, con la scoperta del carillon di campane nella provincia dello Hubei.

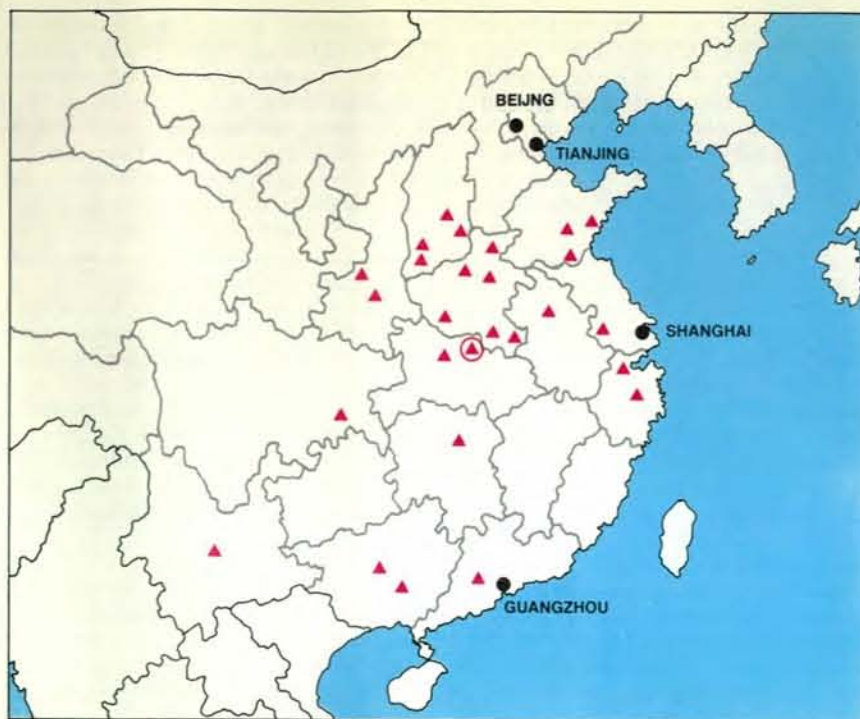
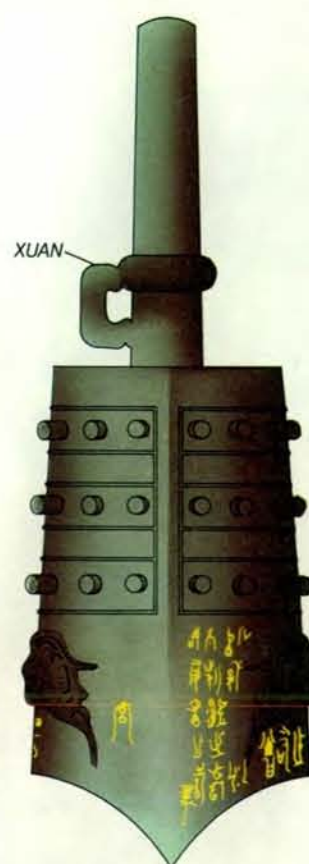
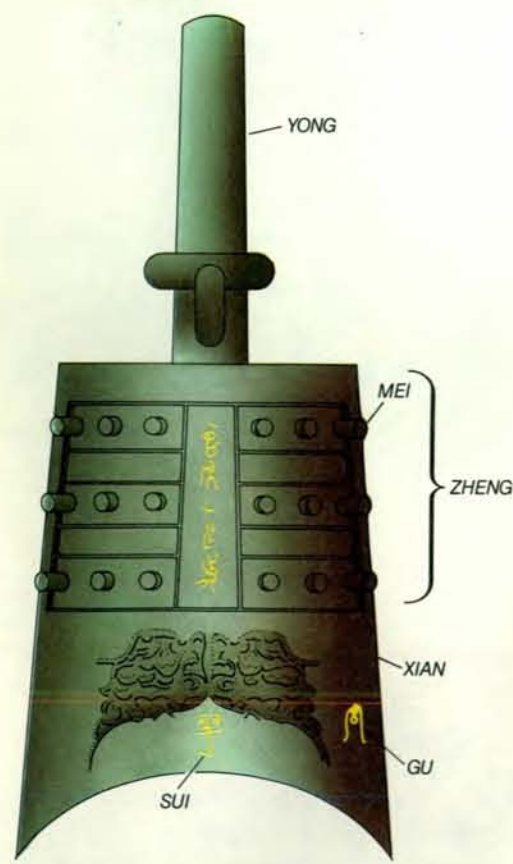
Qualche indizio, in realtà, era disponibile anche prima, ma nessuno vi aveva fatto caso. Studiando il carillon di campane di Jing-li, che era stato riportato alla luce nel 1957 nella provincia dello Henan, i ricercatori dell'Istituto nazionale cinese di ricerche musicali cercarono di eseguire «L'oriente è rosso» usando note ottenute percuotendo le campane al loro centro. Mancava però un Mi4 diesis. Desiderosi di completare il brano, trovarono la nota percuotendo lateralmente una campana Do4 diesis. Il loro successo fu considerato casuale.

Nel 1977 Huang Xiang-peng, Lu Ji, Wang Xiang, Gu Bo-bao e colleghi dell'Istituto esaminarono un carillon di campane scoperto nella provincia dello Shanxi e trovarono che ogni campana, percossa lateralmente, produceva una nota più alta di quella prodotta dalla percussione in posizione centrale. L'intervallo fra le due note era sempre di una terza minore o maggiore, una differenza di frequenza pari rispettivamente a quella fra quattro o cinque tasti consecutivi sulla tastiera di un pianoforte. L'osservazione del gruppo di ricercatori destò interesse e si cominciò a discutere animatamente se il fenomeno fosse casuale o deliberato e se la seconda nota fosse una fondamentale o un sovratono.

I ricercatori si dedicarono allora allo studio di oltre 200 campane dei periodi Shang (XVI-XI secolo a.C.) e Zhou (XI secolo-221 a.C.), coprendo tutto l'arco della produzione storica di carillon di campane. Ne conclusero che le campane erano state costruite deliberatamente in modo da produrre due note distinte. In un carillon le posizioni di percussione laterali erano decorate con glifi della fenice, probabilmente in omaggio ad antiche leggende in cui il canto della fenice connota la musica.

Quando, un anno dopo, nella provincia dello Hubei si scoprì il magnifico carillon di 65 campane, la conclusione dei ricercatori fu definitivamente confermata. Il carillon era stato sepolto nella tomba di Zenghou Yi, sovrano di un antico principato chiamato Zeng, e faceva parte di due grandi orchestre, a loro volta conservate nella tomba. Le campane di bronzo *zhong*, che combinate costituiscono un tipo di carillon chiamato *bian-zhong*, erano intatte e quasi perfettamente accordate. I tesori di Zenghou Yi hanno confermato il progetto a doppia nota delle campane da carillon, grazie a una documentazione dettagliata incisa sulle campane stesse.

Le campane da carillon sono strumenti pensati per l'esecuzione in gruppo, all'interno di orchestre, e di conseguenza le loro caratteristiche debbono essere un po' diverse da quelle di una campana da cerimonia o di una campana da chie-

Il carillon di campane di Zenghou Yi, sovrano di un principato cinese del V secolo a.C., è il più imponente insieme di campane *zhong* trovato fino a oggi. È costituito da tre file di campane montate su un telaio a forma di L. Studi di acustica hanno dimostrato come caratteristiche strutturali speciali permettano a ciascuna campana di produrre due frequenze fondamentali.

I punti di ritrovamento di carillon di campane in Cina, come è indicato nella cartina, sono localizzati soltanto nelle province orientali e sudoccidentali. Dal 1900 sono stati recuperati più di 50 gruppi di campane. Un cerchio identifica l'origine della collezione di Zenghou Yi.

Le caratteristiche delle campane *zhong* sono in rapporto con qualità timbrica e prestazioni della campana, riprodotta in vista frontale (*a sinistra*) e laterale. La campana produce due frequenze perché la sua sezione trasversale è asimmetrica; le note sono perfezionate dai bottoni di bronzo *mei*. Le iscrizioni su alcune campane indicano le posizioni da colpire per ottenere le due note.

sa. Ogni elemento del carillon deve avere un'ampia gamma dinamica e, per consentire l'esecuzione di melodie complesse, le sue note debbono risultare piacevoli all'orecchio ed estinguersi rapidamente, senza code prolungate.

La configurazione geometrica delle campane da carillon è fondamentale per l'ottenimento di queste proprietà acustiche. La campana *zhong* ha una costruzione asimmetrica. A differenza della campana da chiesa, che ha sezione trasversale circolare, una campana *zhong* è schiacciata ai poli: la sua sezione trasversale orizzontale è un ovale appiattito. Il labbro della campana non sta in un piano, ma si inarca verso l'alto nella parte anteriore e posteriore, per poi scendere e formare un piede in foggia di corno sui lati sinistro e destro. Le due facce anteriore e posteriore si incontrano in una costa che prende il nome di *xian*. Nella sezione superiore del corpo della campana, quattro aree sono coperte da 36 bottoni di bronzo, chiamati *mei*.

In un carillon *bian-zhong*, ciascuna campana *zhong* è appesa, tramite un anello chiamato *xuan*, a un gancio montato sulla trave del telaio del carillon. Le campane, prive di battaglio, sono disposte in file e vengono suonate con tipi diversi di battenti. Di solito le campane dei registri superiore e centrale sono appese a livello degli occhi o più in alto e l'esecutore le percuote con mazzuoli a forma di martello. Le campane del registro inferiore sono appese nella fila più bassa e vengono suonate con bacchette, usate per percuoterle con una traiettoria quasi orizzontale.

Un carillon formato da numerose decine di campane richiede da cinque a sette esecutori. Gli esecutori per i diversi registri stanno in piedi su lati opposti del carillon. La bocca della campana è disposta a un angolo di circa 30 gradi rispetto alla verticale, tenuta in equilibrio da un lungo e pesante «naso», chiamato *yong*, che si inclina indietro, in modo che la parte da percuotere fronteggi l'esecutore. A seconda dell'altezza della nota che producono, le campane *zhong* hanno un'altezza che va da pochi centimetri a parecchie decine di centimetri: le maggiori sono strumenti relativamente grandi per gli standard moderni.

Le due frequenze fondamentali che hanno dato alle campane la loro fama hanno origine da due regioni distinte sulla faccia dello strumento. Una ha origine dalla posizione *sui*, nella zona centrale bassa della campana; l'altra ha origine dalle posizioni *gu*, che si trovano a sinistra e a destra della *sui*. Si dice che il termine *sui* significhi «specchio», probabilmente in riferimento alla somiglianza fra la curva del labbro della campana e lo specchio ustorio concavo usato a quell'epoca per accendere il fuoco. *Gu* significa «tamburo» o «che produce musica». A causa di una forte predilezione culturale per la destra, la posizione *gu* di destra veniva percossa più spesso della sinistra.

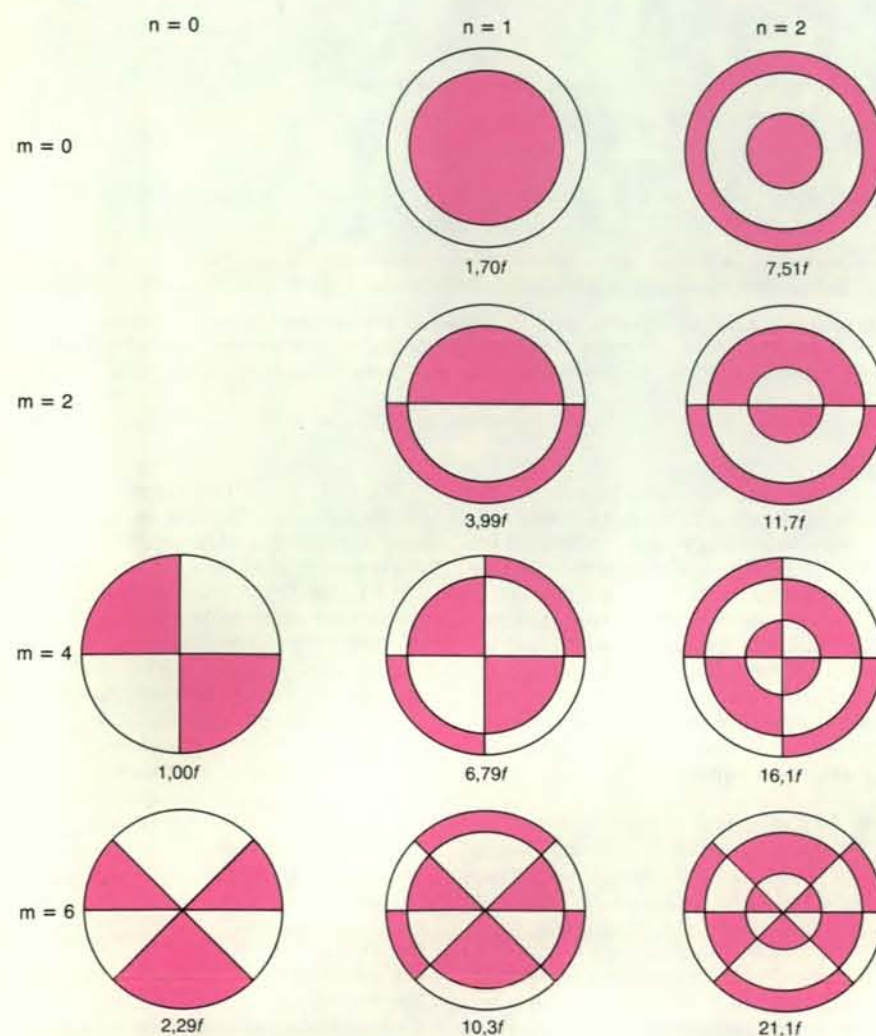
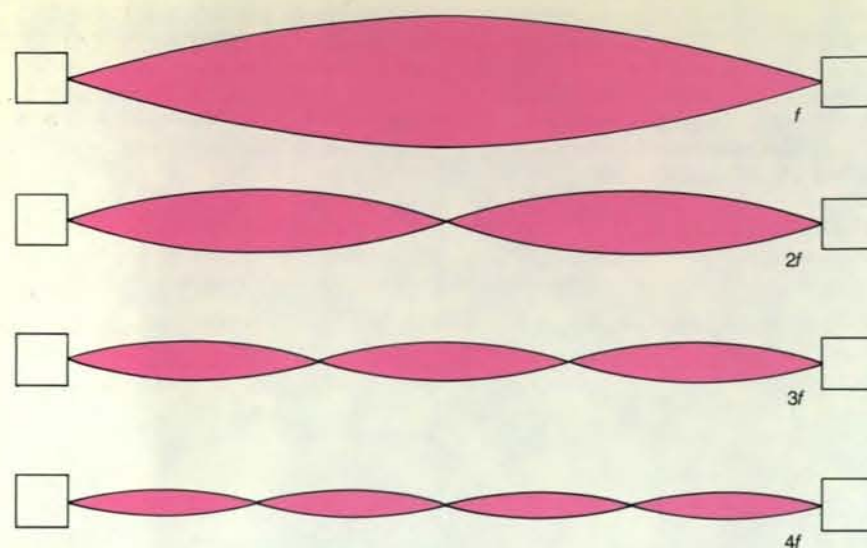
L'analisi acustica ha dimostrato la precisione della campana *zhong*. I progettisti zhou avevano imparato a manipolare non solo le due fondamentali del suono delle campane, ma anche le parziali superiori. In fotografia, Wang Xiang (*seduto*) e Huang Xiang-peng dell'Istituto nazionale cinese di ricerca musicale determinano le note del carillon di Zenghou Yi. Nonostante l'età, le campane sono intonate quasi alla perfezione.

Nella musica di gruppo, le campane a due note permettono prestazioni migliori e riducono le dimensioni complessive del carillon. Il disegno delle campane occidentali non è altrettanto pratico. La nota che si percepisce alla percussione di una campana rotonda non è la fondamentale, ma la nota che sta un'ottava sopra la fondamentale. Per esempio, una campana americana moderna che ha un Do centrale (Do3 a 256 cicli al secondo) come fondamentale produce una nota percepita Do4 (a 512 cicli al secondo). Il peso di questa campana è di 800 libbre, circa 400 chilogrammi. In ossequio a una regola approssimata, la frequenza della nota prodotta da una campana rotonda è inversamente proporzionale alla radice cubica della sua massa: un dimezzamento della frequenza, quindi, richiederebbe una campana con una massa otto volte superiore. Una campana americana che dia un Do centrale come nota percepita, pertanto, peserebbe oltre tre tonnellate, circa 500 chilogrammi più di tutte le campane e i telai del carillon di Zenghou Yi messi insieme.

L'incapacità di produrre una campana che possa eseguire la sua fondamentale acustica ha avuto conseguenze pesanti relativamente ai materiali richiesti nella costruzione di campane e ai vincoli imposti alla fusione. Il progetto della campana da chiesa, tuttavia, è il risultato di secoli di sperimentazione attenta. Il progetto cinese, risultato di migliaia di anni di attività artigianali, è andato perduto dopo il periodo Han (206 a.C.-220 d.C.). Quali segreti dell'acustica incorporava, che sono invece sfuggiti ai progettisti occidentali?

Una campana è un corpo acustico molto complesso. Le sue parziali non possono essere formulate come rapporti aritmetici semplici, a differenza delle parziali di una corda completamente elastica o di una colonna d'aria in vibrazione, le cui frequenze corrispondono ai rapporti 1 : 2 : 3 : 4 : 5 : 6 e così via. Sia la campana *zhong* schiacciata ai poli, sia la campana da chiesa a bocca rotonda sono adattamenti particolari del sistema acustico delle lamine vibranti. Per lamine e campane, l'aumento dello spessore e dell'elasticità del materiale vibrante aumentano la frequenza di vibrazione, mentre la frequenza di vibrazione diminuisce al crescere del diametro e della densità.

Mary D. Waller della London School of Medicine for Women ha studiato i modi normali di lamine circolari in vibrazione e ha pubblicato i suoi risultati nel 1937. Le sue figure nodali sono costituite fondamentalmente da raggi (indicati con m) distribuiti simmetricamente attorno al centro della lamina, e da circonferenze (indicate con n) concentriche alla lamina. Il modo più semplice, quello che corrisponde alla frequenza fondamentale, presenta quattro linee nodali che dividono la lamina in quattro sezioni vibranti, dalla forma a fetta di torta. I settori adiacenti si muovono in ogni istante in direzioni opposte. Il modo successivo, che genera una frequenza pari a 1,7 volte quella fondamentale, ha solo una circonferenza nodale, che delinea un cerchio più piccolo interno e una corona circolare esterna. Altri modi nascono da ulteriori combinazioni di raggi e circonferenze, creando una ricca interazione di frequenze parziali.

Dal punto di vista acustico, una campana di foggia rotonda si comporta come una lamina stirata fino ad assumere una forma di campana e appesa al centro. I modi normali di una campana rotonda, vista dall'alto, sono del tutto simili alle figure che si vedono sulle lamine circo-

Le frequenze parziali, che costituiscono la qualità timbrica di un suono, sono il risultato dei diversi modi di vibrazione di un corpo. La frequenza più bassa è detta fondamentale; le altre frequenze sono parziali superiori o sovratoni. Quando si pizzica una corda ben tesa (*in alto*), questa produce parziali la cui frequenza è un multiplo intero della frequenza *f* della fondamentale. Una lamina vibrante (*in basso*), invece, produce parziali che non stanno fra loro in rapporti aritmetici semplici. Le frequenze dipendono dalla combinazione di raggi nodali (*m*) e circonferenze nodali (*n*), regioni della lamina che rimangono immobili mentre le altre parti vibrano. Il colore e il bianco rappresentano movimenti in direzioni opposte. Una campana, indipendentemente dal fatto che sia rotonda o schiacciata ai poli, è un tipo particolare di lamina vibrante.

lari. Il moto vibratorio è più intenso sul bordo. Quando la campana viene percossa, la faccia colpita viene spinta verso l'interno e le regioni adiacenti sono spinte verso l'esterno. Il bordo quindi passa attraverso la sua forma circolare iniziale per poi costituire un'altra circonferenza oblunga in senso perpendicolare rispetto alla prima.

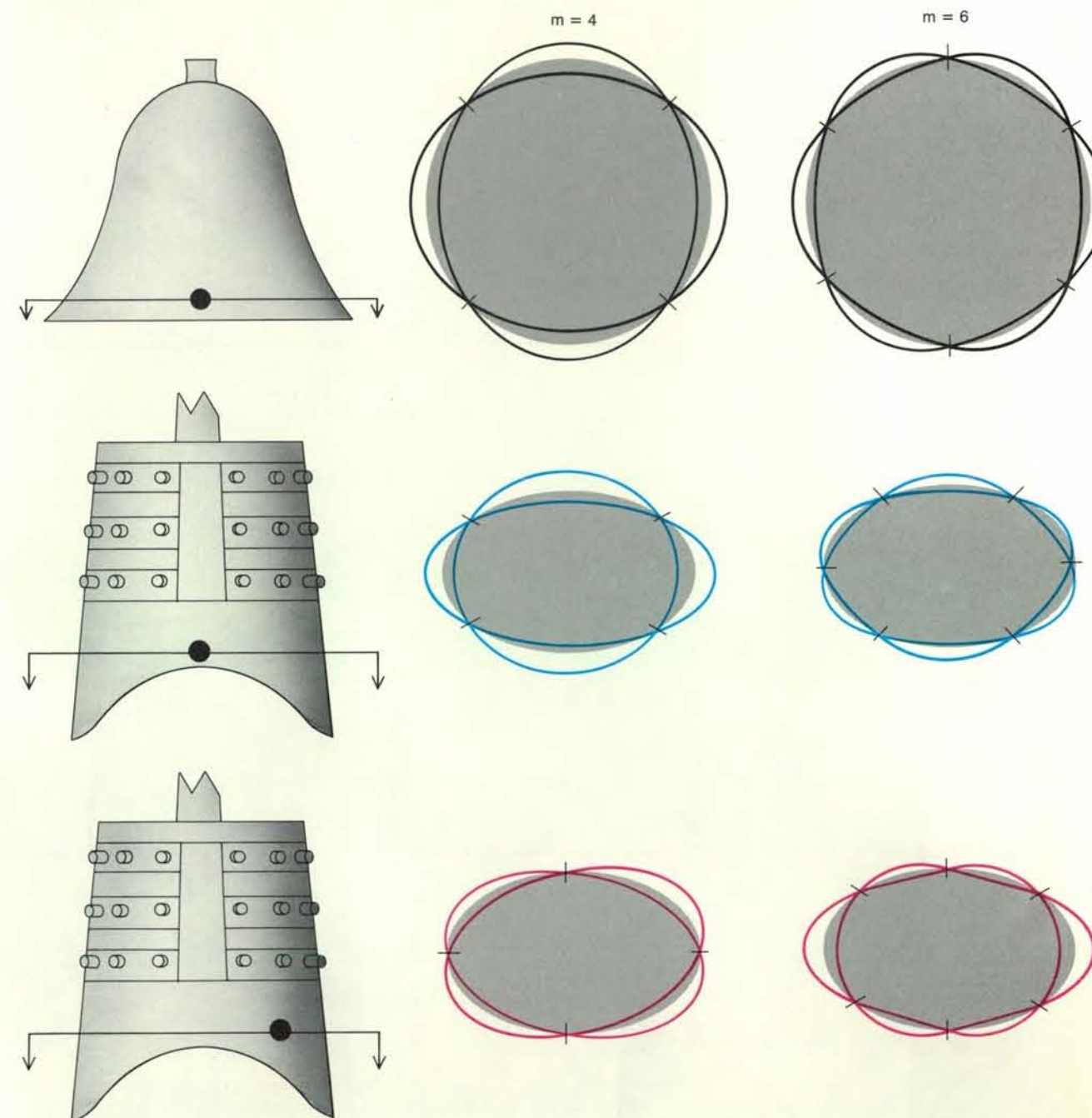
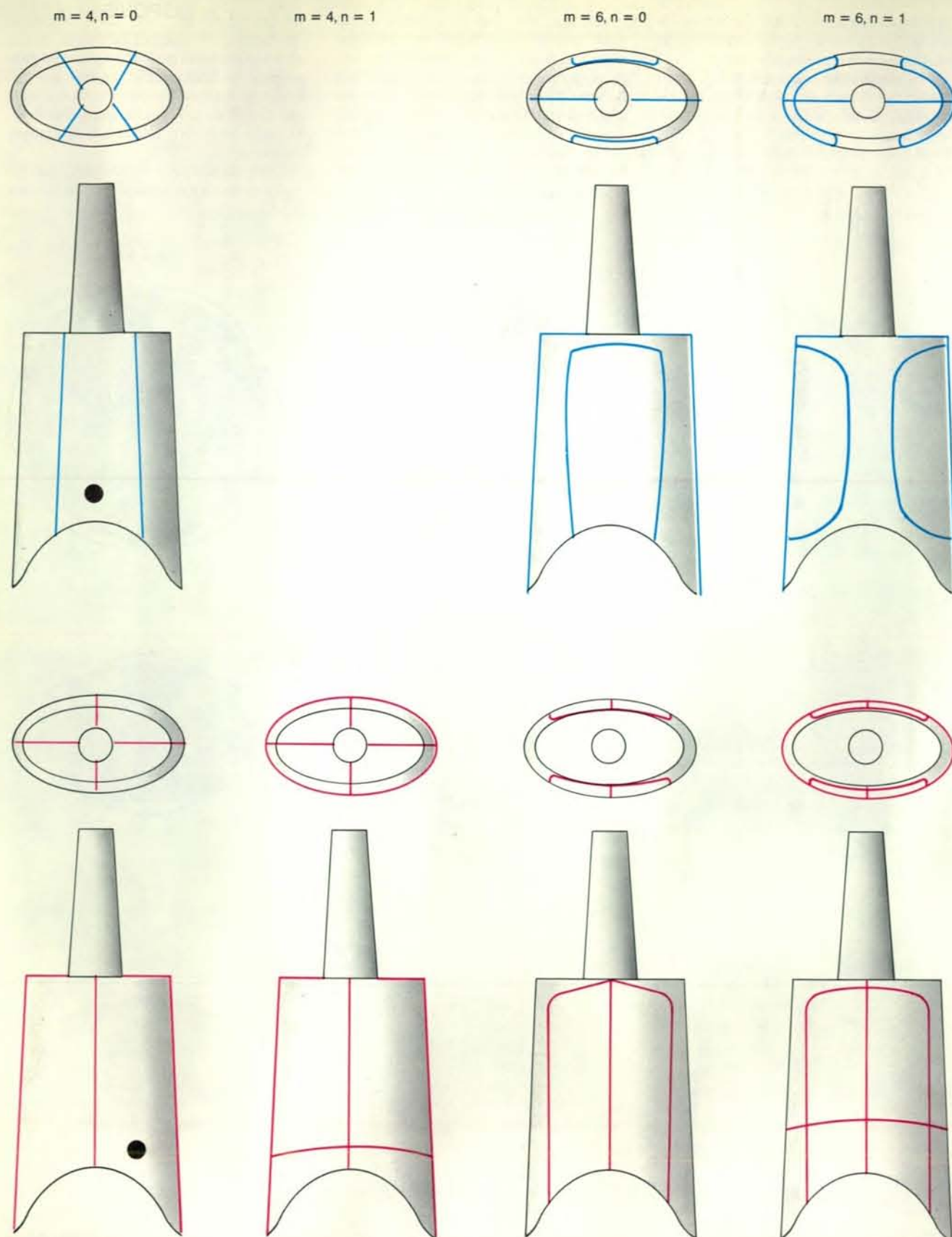
Nel corso di queste vibrazioni certe parti della campana rimangono relativamente immobili rispetto ad altre, nella fattispecie i punti nei quali si intersecano le diverse distorsioni circolari. Questi punti rappresentano le linee nodali della campana, chiamate meridiani nodali. Non si tratta tuttavia di nodi in senso stretto, perché le intersezioni non sono esatte e quindi si ha ancora qualche movimento nel piano della superficie. Questo movimento è quello che fa «suonare» un calice di vetro quando si fa scorrere un dito umido attorno al suo bordo.

Su una campana rotonda, i meridiani nodali sono distribuiti uniformemente, come i raggi nodali di una lamina. Grazie a questa simmetria, con una campana rotonda può essere usato un battaglio, poiché si ottengono sempre gli stessi effetti di vibrazione qualunque sia il punto della superficie che viene percosso. I meridiani della campana asimmetrica *zhong*, invece, non sono distribuiti uniformemente: da questa asimmetria consegue che, per qualunque numero dato di linee nodali, è possibile più di una disposizione spaziale.

In effetti, la campana *zhong* ha due insiemi ben definiti di modi, che possono essere attivati selettivamente percuotendo posizioni diverse sulla campana: le posizioni *sui* e *gu*. In corrispondenza delle due frequenze fondamentali i modi hanno lo stesso numero di linee nodali ($m = 4$ e $n = 0$), ma la collocazione delle linee è diversa, e quindi le frequenze generate sono diverse. Questa «degenerazione» di modi vibrazionali spiega le straordinarie caratteristiche acustiche delle campane da carillon. In generale il modo *gu* produce frequenze più alte del modo *sui*, ma i due modi non si escludono a vicenda: hanno in comune certe parziali ad alta frequenza che nascono dai modi più complessi.

Gli antichi cinesi perfezionarono il progetto delle campane a due note in modo da rendere i due insiemi di modi distinti per il loro carattere acustico, ma paragonabili per la loro funzione musicale. Essi lavorarono per separare le due frequenze di ciascuna campana escludendo le qualità comuni. Per esempio, quando viene colpita la posizione *sui*, le facce e i lati della campana subiscono i movimenti maggiori, mentre le aree *gu* rappresentano i nodi a riposo, che non producono suono. Le aree *sui* diventano nodi quando la campana viene percossa in una posizione *gu*. Quindi ciascuna posizione di percussione è situata nel punto che viene meno sollecitato quando viene percossa l'altra posizione, che è dunque anche l'area meno significativa per la produzione della nota alternativa.

Per localizzare con tanta precisione i meridiani nodali dei due modi fondamentali, gli antichi cinesi dovevano possedere una comprensione teorica della fisica della musica di gran lunga superiore a quella attribuita loro inizialmente dagli storici. Considerata una tale comprensione, sarebbe abbastanza immediato, anche se non facile, trovare le posizioni di percussione che isolano le fondamentali. Per ottenere la migliore risoluzione fra le frequenze, però, debbono essere separate non solo le fondamentali, ma anche le parziali superiori. I meridiani nodali per i modi importanti delle parziali *gu* convergono naturalmente nella posizione *sui*; pertanto non danno un contributo alla nota *sui*. I meridiani dei modi delle parziali *sui*, però, non si raccolgono nella posizione *gu*, e pertanto qualche traccia delle parziali *sui* può interferire quando viene percossa la nota *gu*. Qui è da cercare la spiegazione razionale del labbro arcuato della campana *zhong*.

Quando abbiamo esaminato i due insiemi di modi normali, abbiamo trovato

Il movimento del bordo di una campana, in seguito alla percussione dello strumento, illustra la distribuzione delle linee nodali per una campana rotonda (*in alto*) e per le due note di una campana cinese *zhong*, note chiamate *sui* (*al centro*) e *gu* (*in basso*). Il grigio indica le posizioni di riposo; le linee in colore e in nero indicano i cambiamenti di forma della campana dopo la percussione. Sono rappresentati i modi in cui *m* è uguale a 4 e a 6. In questi modi i raggi nodali di una campana rotonda sono distribuiti uniformemente; di conseguenza è possibile un unico schema di distribuzione. Poiché invece una campana da carillon è di forma schiacciata ai poli, un dato numero di linee nodali può essere disposto in molti modi diversi e figure nodali diverse generano frequenze diverse. Come si può vedere, la figura dei raggi nodali è determinata dal punto in cui la campana è stata colpita. I progettisti cinesi riuscivano a rendere massima la separazione fra le due note della campana *zhong* facendo in modo che le linee nodali di una nota fungessero da posizione di percussione per l'altra.

Le figure nodali per le note *sui* (*in alto*) e *gu* (*in basso*) descrivono i modi delle fondamentali e delle parziali superiori più importanti. Gli schemi «in pianta» sopra ogni campana utilizzano il sistema della lamina vibrante per enumerare raggi e circonferenze. Le linee orizzontali valgono come circonferenze anche se non compiono un giro lungo il perimetro della campana, ma vicino alle spalle non contribuiscono al suono e non vengono contate. Come mostra l'illustrazione in alto di pagina 98, la nota *sui* ha due sovratoni dominanti, mentre la nota *gu* ne ha tre.

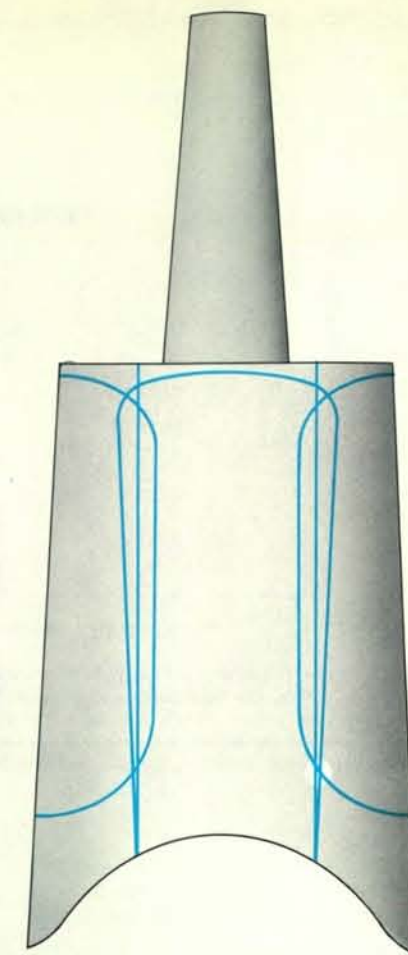
che l'arco nel labbro della campana, modificando la forma della «lamina» vibrante, modifica anche le figure nodali delle più importanti fra le parziali *sui*. Grazie al labbro ad arco, i meridiani delle parziali *sui* convergono nel punto designato, non a caso, come posizione *gu*, tipicamente a tre quinti della distanza che separa la posizione *sui* dalla costa *xian*. La posizione di percussione è così critica per il raggiungimento della frequenza e della qualità timbrica desiderate, che gli antichi decorarono le loro campane con iscrizioni che indicavano in modo privo di ambiguità le posizioni *sui* e *gu*. Né il progetto del bordo concavo, né la precisione nell'identificare la convergenza delle linee nodali potrebbero essere state casuali.

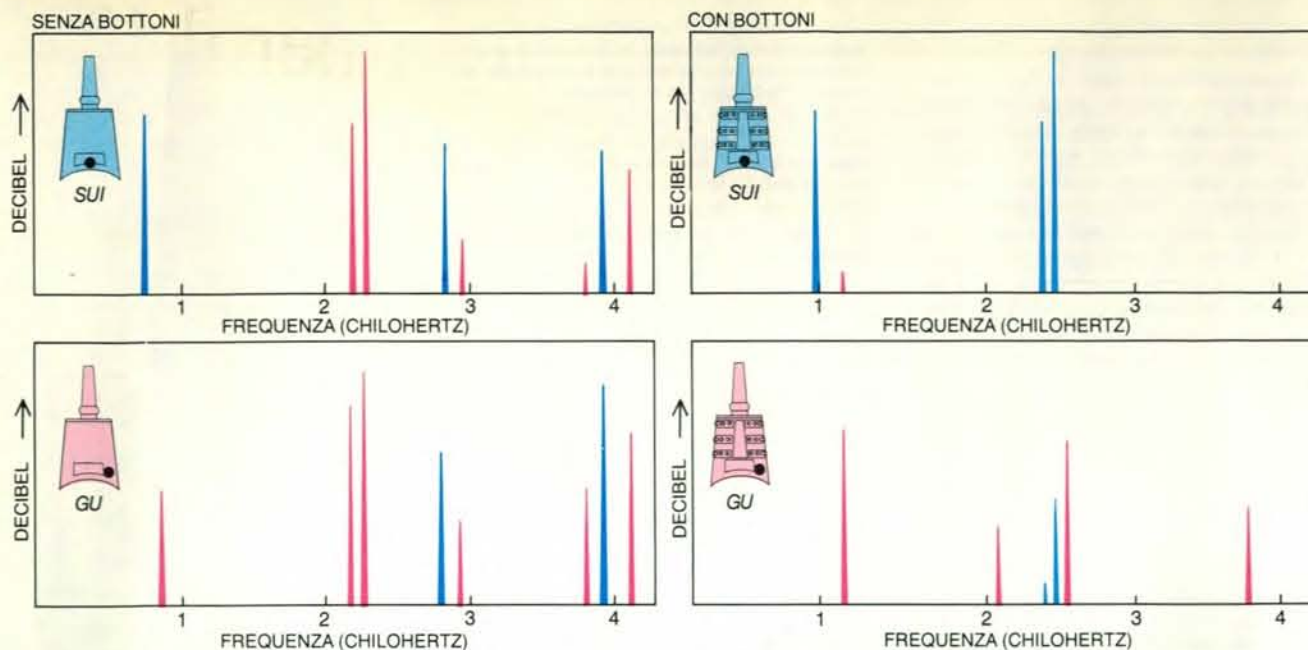
Anche i bottoni *mei* raccolti nella parte alta di tutte le campane, fuorché quelle di registro superiore, non sono semplicemente motivi ornamentali, ma contribuiscono a equilibrare l'intensità delle due fondamentali in modo che i loro volumi siano paragonabili. Cosa ancor più importante, i *mei* fungono da ulteriore dispositivo per la separazione delle due note della campana. Studi recenti condotti in laboratorio hanno permesso di scoprire che i bottoni modificano le strutture complessive delle parziali, ovvero gli spettri di frequenza, delle note *sui* e *gu*. I *mei* forniscono peso ulteriore attorno alle spalle della campana, modificando le figure nodali nella parte superiore dello strumento. Senza i *mei*, le fondamentali *sui* e *gu* si distinguono facilmente, ma hanno in comune talune parziali di alta frequenza. Con i bottoni si ha uno scostamento delle frequenze delle parziali, cosicché la sovrapposizione fra i due insiemi di parziali è minima.

L'intervallo fra le due frequenze fondamentali di una campana *zhong* viene determinato dalla fusione e dall'accordatura. La scelta dell'intervallo è arbitraria, ma deve essere tale da adattarsi alla progressione melodica delle composizioni che si andranno a eseguire con lo strumento. L'intervallo inoltre non deve essere dissonante, poiché, nonostante gli sforzi dei progettisti, qualche traccia della nota secondaria può rimanere dopo l'estinzione del suono primario. I progettisti zhou accordavano le loro campane in modo che gli intervalli delle parziali superiori, così come quelli fra le fondamentali, fossero armonici. La seconda parziale della nota *gu*, per esempio, si trova sempre un'ottava più una terza maggiore o minore sopra la nota *sui*; le parziali di una campana *zhong* le cui fondamentali siano separate da una terza minore stanno fra loro nel rapporto 1 : 1,2 : 2,4 : 2,81 : 3 e quelle di una campana le cui fondamentali siano separate da una terza maggiore sono nel rapporto 1 : 1,25 : 2,5 : 2,81 : 3.

Se si mettono a confronto gli intervalli fra le fondamentali di campane di carillon di periodi diversi, risulta evidente una tendenza a intervalli di terza maggiore e minore, rappresentata nella collezione di campane di Zenghou Yi. In assenza di esempi di musica antica, possiamo solo ipotizzare che questa preferenza corrispondesse a un gusto per le terze maggiori e minori nelle composizioni musicali di quel periodo. In Europa questi intervalli sono stati considerati armoniosi solo nel XII secolo.

Quando è stato ritrovato il *bian-zhong* di Zenghou Yi, le sue campane erano quasi perfettamente accordate, nonostante il lungo periodo di tempo in cui sono rimaste sepolte. Un secondo gruppo di 36 campane scoperto nello stesso principato di Zeng nel 1981 era accordato ancora meglio. Di solito una lamina vibrante viene accordata per aggiunta o rimozione di materiale, ma con questo metodo sembrerebbe impossibile accordare una delle note della campana senza modificare l'altra, dal momento che sono ambedue contenute in un unico corpo continuo. Come facevano i cinesi ad accordare due note su un'unica campana *zhong*?

Anche in questo caso gli antichi sfruttavano la loro notevole abilità nell'identificare le linee nodali. Potevano accordare una singola nota limando bronzo da una superficie interna della campana se, nel limare, seguivano attentamente le linee nodali dell'altra nota. Quindi accordare la nota *sui* comportava togliere me-

La convergenza di linee nodali per le tre parziali dominanti della nota *sui* è una conseguenza importante del progetto delle campane da carillon. Se il labbro della campana fosse piatto, i nodi *sui* non si intersecherebbero mai e un colpo in qualunque parte della campana eciterebbe almeno una frequenza *sui*. Il labbro ad arco della campana *zhong*, però, modifica l'andamento delle linee nodali così che esse si raccolgano nella posizione *gu* (*pallino bianco*). Questa convergenza fa sì che la nota *gu* risulti «pulita» da qualunque eco della nota *sui*.

Ologrammi laser evidenziano i modi fondamentali di vibrazione per le note *sui* (*a sinistra*) e *gu* (*a destra*). Le aree in bianco sono nodi; le linee scure, le zone dove il movimento è intenso. E' evidente la convergenza complementare delle linee nodali. Le immagini sono dovute al modo diverso in cui le superfici in movimento e quelle stazionarie della campana riflettono la luce.

I grafici delle frequenze parziali evidenziano l'importanza dei bottoni *mei* per la separazione delle note *sui* (*in blu*) e *gu* (*in rosso*). Le frequenze generate dalle campane senza *mei* (*a sinistra*) sono pressoché identiche per le due posizioni: solo le fondamentali sono molto diverse. Con l'aggiunta dei bottoni (*a destra*), le figure nodali nella parte superiore della campana si dispongono in modo diverso, modificando alcune frequenze ed eliminandone altre. Dal profilo della nota *sui* scompaiono quasi tutte le tracce di suoni *gu*, e nel suono *gu* resta solo qualche elemento delle parziali *sui*. I bottoni *mei* servono anche ad aumentare l'intensità della fondamentale *gu*, in modo che corrisponda a quella della *sui*.

tallo dalle linee nodali *gu* e viceversa. Inoltre, molte campane erano fuse con tanta precisione da non richiedere alcuna modifica.

La metodologia della fusione in bronzo con stampi a conchiglia di ceramica era perfettamente sviluppata nell'antica Cina. Anche così, però, la produzione di una campana *zhong* di registro basso rappresentava un compito formidabile. Spesso oggetti grandi e complessi venivano ottenuti fondendo le parti separatamente, per poi unire le componenti in uno stampo finale; ma la campana da carillon in bronzo, non importa quanto grande, era sempre fusa in un pezzo unico.

Il carillon occidentale moderno è costituito da un gruppo di campane attentamente selezionate in modo da dare intervalli cromatici ben temperati. Le campane vengono suonate da una tastiera. Poiché alcune parziali formano dissonanze, il carillon può essere usato per generare solo un numero limitato di accordi, tanto che di solito suona un'unica melodia. Per ottenere effetti soddisfacenti sono necessarie conoscenza e abilità notevoli; anche installazioni molto costose possono dare prestazioni deludenti. Se il carillon è il meglio che il mondo moderno abbia da offrire, ci si può chiedere come abbiano fatto i cinesi a sviluppare il progetto del *bian-zhong* tanto tempo fa.

Sin dai tempi antichi, i musicisti cinesi sono stati sensibili a differenze sottili nella qualità timbrica; questa sensibilità si è manifestata in molti dei loro strumenti. Una volta, per esempio, chiesi a un fiatista cinese la spiegazione razionale del progetto dello *shuang-guan*, uno strumento costituito da due oboi cilindrici, apparentemente identici. Mi rispose: «Non esistono due tubi dotati di ancia che abbiano lo stesso spettro timbrico. Con due tubi l'esecutore controlla una gamma timbrica più estesa e può essere più selettivo nei suoi timbri».

Forse questa acuta sensibilità per la struttura timbrica portò gli antichi cinesi a sperimentare un progetto di campana leggermente ellittico, che dà origine a un insieme più completo di modi vibrazionali e a una gamma più ampia di possibili sovratoni. Le campane *ling*, tipiche del primo periodo Shang, erano schiacciate ai poli, ma avevano ancora il battaglio e producevano un solo suono; le loro qualità musicali possono essere dedotte dal fatto che erano usate come campanelle per cani e mucche.

Nei tentativi successivi, i cinesi debbono essersi trovati alle prese con campane schiacciate ai poli che producevano due frequenze distintamente diverse, ma con timbri solo strani e confusi. La campana a mano *zheng*, grande e schiacciata ai poli, che precede la campana da carillon di centinaia di anni, esemplifica questo stadio di sviluppo. In effetti, le campane *zheng* hanno dato il loro nome alla metà superiore della campana *zhong*, e lo *yong* che tiene la campana *zhong* inclinata a 30 gradi è derivato dal manico *zheng*. La campana *zheng* senza battaglio, però, ha un suono *gu* confuso, perché la sua bocca appiattita non determina la convergenza delle linee nodali *sui*.

Grazie alle modifiche apportate alla superficie, alla sezione trasversale e allo spessore della campana, le due frequenze sono state portate progressivamente ad assumere una relazione armonica ben definita e le loro singole note sono state precisate; alla fine ne è scaturito il raffinato disegno *zhong*. Della stessa epoca delle campane *zhong* sono la *niu-zhong*, anch'essa a due note, e la *bo* cerimoniale, la quale non possiede l'agilità musicale della campana *zhong*. Ambedue questi tipi di campane venivano suonati in posizione verticale.

L'arte della musica ha una storia lunga e ricca, parallela all'ascesa della civiltà umana. I principi racchiusi nel carillon di campane di Zenghou Yi fanno pensare che la scienza della musica possa avere una storia quasi altrettanto lunga, e sicuramente altrettanto ricca. La Cina dei periodi Shang e Zhou aveva sviluppato la scienza acustica al livello essenziale per sostenere la complessa arte musicale di quell'epoca. Fisica e ingegneria collaboravano al perfezionamento di strumenti a fiato, a corda e a percussione, e a favorire la loro fusione orchestrale. Sono tutti risultati che hanno corrispondenti moderni; il progetto complessivo di un grande gruppo di campane a due note per l'esecuzione musicale, invece, è un risultato che non ha l'eguale nella moderna fisica della musica.

La successione storica dei modelli di campane suggerisce la strada seguita dagli antichi cinesi per arrivare al modello *zhong*. La *ling*, risalente a oltre 3600 anni fa, aveva un battaglio e produceva solo una nota poco significativa. Il campanello a mano *zheng*, un esempio primitivo di progetto a doppia nota, era usato dai soldati cinesi che lo portavano con la bocca rivolta verso l'alto. Anche la campana *niu-zhong* produce due note, ma le manca il timbro chiaro della *yong-zhong* nel registro più basso. Alcune *niu-zhong* sono visibili nella fila superiore del carillon di pagina 93. Un'altra campana della stessa epoca della *zhong*, la *bo*, funge-va da campana cerimoniale a una sola nota. I disegni non sono in scala.

La più antica tecnologia

Ricostruendo una tecnica di scheggiatura utilizzata dai nostri antenati dell'Età della pietra oltre due milioni di anni fa, un giovane archeologo ha contribuito a chiarire le modalità d'uso dei primi manufatti litici

di Nicholas Toth

Gli esseri umani non sono gli unici animali che usano arnesi. Numerose specie di mammiferi e di altri gruppi animali utilizzano materiali come pietre o rami nel procurarsi il cibo o in altre attività. I nostri parenti più prossimi, gli scimpanzé, di quando in quando modificano elementi naturali per formare arnesi più o meno semplici, destinati a raccogliere insetti commestibili o acqua, o a praticare la pulitura del pelo. Non solo, ma il comportamento è appreso: i giovani scimpanzé diventano via via più esperti imitando gli anziani. Anche fra gli scimpanzé, però, l'uso di arnesi non sembra essere indispensabile, mentre per gli esseri umani di oggi esso costituisce un mezzo di sopravvivenza insostituibile. Pochi di noi potrebbero sopravvivere in una zona selvaggia senza disporre di arnesi di sorta. Come si è giunti a ciò?

La risposta è racchiusa nel nostro passato africano compreso fra due e tre milioni di anni fa. In qualche momento di quel periodo i nostri antenati intrapresero un modo di vita caratterizzato da un fatto nuovo e cruciale mai prima osservato in altre specie: la fabbricazione di arnesi mediante una tecnologia di scheggiatura della pietra. Questa novità tecnologica sembra essersi sviluppata parallelamente a mutamenti non meno cruciali dell'alimentazione e della forma biologica. Come i primi manufatti di pietra si inserissero nella vita dei loro costruttori non è del tutto chiaro. Gli studi tradizionali sono stati ostacolati dalla mancanza di informazioni che potessero essere utilizzate in modo comparativo, dal momento che i più antichi ominidi (cioè i rappresentanti della famiglia zoologica che comprende la specie umana attuale) non erano esattamente uguali all'umanità di oggi, né agli scimpanzé o ai babbuini.

Anche per fare fronte a questa carenza, alcuni archeologi hanno cominciato a mettersi nei panni dei più antichi ominidi per cercare di capire come i primi arnesi fossero fatti e adoperati. Nella località di Koobi Fora, nel Kenya settentrionale, dove i più antichi resti archeologici risalgono a un milione e 900 000 anni fa, io stesso ho svolto una serie di esperimenti replicativi e manuali basati sulla fabbricazione e sull'uso dei tipi di arnesi più primitivi. I risultati di questa «archeologia sperimentale» sono affascinanti. Il lavoro da me fatto suggerisce che la più antica tecnologia della pietra desse grande importanza alle schegge, e non soltanto al «nucleo» scheggiato, ossia la pietra di partenza, come abitualmente si pensa. Risulta anche evidente che già intorno a due milioni di anni fa i nostri antenati erano diventati destrimani, una scoperta che può avere importanti implicazioni per capire l'evoluzione del cervello umano.

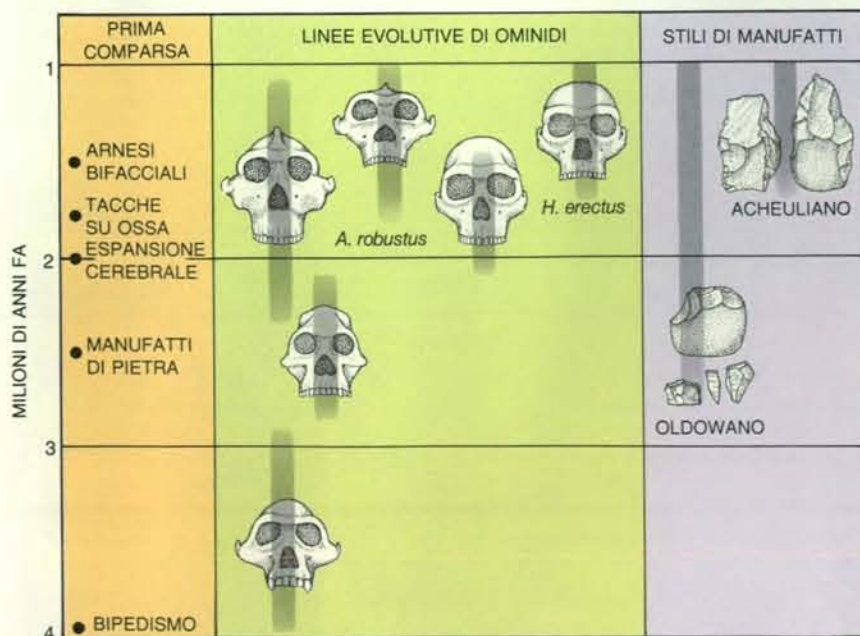
L'archeologia sperimentale ha avuto una storia lunga, ma abbastanza disordinata. A partire dal XIX secolo, non pochi antiquari si misero a scheggiare la pietra allo scopo di scoprire come gli arnesi preistorici fossero stati fabbricati. Purtroppo alcuni dei migliori archeologi «sperimentali» dello scorso secolo non erano che imbroglioni, i quali fabbricavano arnesi di pietra per venderli: una pratica che non contribuì al buon nome degli esperimenti. In generale né gli esercizi legittimi né quelli illegittimi furono particolarmente sistematici, e con poche eccezioni l'idea di un'archeologia che includesse esperimenti funzionali approfonditi era destinata a restare lettera morta fino alla seconda metà del XX secolo. In quel momento, alcuni pionieri dell'archeologia preistorica africana, come J. Desmond Clark dell'Università della California a Berkeley e Louis Leakey, iniziarono a porre l'impostazione sperimentale nello studio della litotecnica su una base più rigorosa, indirizzata a precisi problemi.

Ciò nonostante la maggior parte del lavoro recente si è limitata a sfiorare la superficie delle possibilità di impostazione sperimentale. Molti esperimenti, per fare un esempio, si sono concentrati sulla fabbricazione di un determinato tipo di arnese per vedere come esso fosse ottenuto. Molti di tali arnesi erano relativamente sofisticati a paragone dei tipi più primitivi. Clark e altri studiosi, invece, hanno affidato gli arnesi preistorici a cacciatori-raccoglitori africani contemporanei, nel tentativo di scoprire come i manufatti antichi potessero essere stati usati in origine. Contrariamente a questi orientamenti, io ho deciso di fabbricare e di usare l'intera gamma di manufatti corrispondenti alle fasi più primitive della tecnologia. In tal modo speravo di ricostruire non tanto le singole tecniche, ma l'universo complessivo delle possibilità tecnologiche a disposizione dei nostri primi antenati.

Koobi Fora si è dimostrato un luogo eccellente per il tentativo. La zona si trova sulla sponda nord-orientale del lago Turkana, nel Kenya settentrionale, ed è l'area di un programma di ricerca diretto dal compianto Glynn Isaac (allora all'Università della California a Berkeley) e da Richard E. F. Leakey dei Musei Nazionali del Kenya. Sotto la loro direzione, un gruppo internazionale di ricercatori, comprendente paleontologi, anatomisti, geologi, geochimici e archeologi, è venuto svolgendo un programma coordinato di indagini a partire dal 1969. In uno specifico progetto, condotto da Isaac e da John W. Harris dell'Università del Wisconsin a Milwaukee, si sono scavati i siti preistorici della zona di Koobi Fora. Le moderne tecniche utilizzate hanno permesso il recupero esauriente di tutte le categorie di materiali archeologici (si veda l'articolo *La spartizione del cibo negli ominidi protoumani* di Glynn L. Isaac in «Le Scienze» n. 118, giugno 1978).

Da un tale bottino scientifico c'è molto da imparare, se si considera che Koobi Fora offre 20 siti scavati dell'inizio del-

L'autore macella un elefante con arnesi di pietra da lui stesso fabbricati, mentre un collaboratore prepara la carne per pesarla. (Questo grande elefante africano era morto per cause naturali: nessuno degli animali usati dall'autore nelle sue ricerche è stato ucciso in vista dell'esperimento.) L'uso simultaneo di schegge e di arnesi più pesanti, in questo esperimento di macellazione, è caratteristico delle tecnologie litiche più primitive che si svilupparono in Africa circa due milioni e mezzo di anni fa.

In Africa si sono svolti molti avvenimenti importanti all'inizio dell'evoluzione umana. La cartina illustra i siti dell'Età della pietra ritenuti più antichi di un milione di anni. Koobi Fora è la località in cui l'autore ha effettuato gran parte della sua «archeologia sperimentale».

Lo sviluppo del cervello fu senza dubbio in relazione con la nascita della tecnologia. Per più di un milione di anni, all'epoca dei primi manufatti di pietra, esistettero due linee di ominidi. Una, quella degli australopiteci più recenti (fra cui *Australopithecus boisei*), aveva probabilmente una dieta basata su vegetali coriacei. L'altra linea, dotata di cervello più grande e apparentemente di una dieta di migliore qualità, comprendeva *Homo habilis*, un antenato dell'uomo attuale. La più antica tecnologia della pietra, detta oldowana, si basava su manufatti ottenuti da ciottoli. A partire da un milione e mezzo di anni fa, fece la sua comparsa un nuovo «indirizzo» tecnologico, quello acheuliano, con utensili ricavati da grandi schegge, talvolta lunghe più di 20 centimetri.

l'Età della pietra, databili all'arco di tempo tra un milione e 900 000 e un milione e 400 000 anni fa. I siti si trovano in una moltitudine di situazioni geografiche e sedimentarie, atte a fornire una quantità di notizie comparate. Oltre a ciò, tutti i tipi di roccia utilizzati nella preistoria per la confezione di arnesi sono tuttora reperibili nei letti dei torrenti, e sono quindi a disposizione dello sperimentatore. Il materiale principale che la natura offriva agli antichi ominidi era una roccia lavica di grana media. Altre rocce, usate con minore frequenza, erano la selce, il quarzo e l'ignimbrite (una roccia vulcanica che si forma a partire da colate di ceneri ardenti).

I siti di Koobi Fora si trovano oggi all'interno di un'area erosa semiarida, circondata di rilievi vulcanici a est e chiusa dal lago Turkana a ovest. Ma tra 1,9 e 1,4 milioni di anni fa la zona appariva ecologicamente molto diversa. Innanzi tutto, essa era molto più ricca: c'era una piana rivestita di savana-prateria, mentre lungo i corsi d'acqua che dalle alture scendevano verso il precursore del lago Turkana c'erano foreste a galleria, esili strisce verdi che correvano parallele su ciascun lato degli alvei. Il paesaggio appariva popolato di animali, fra i quali molte forme attualmente estinte di elefanti, rinoceronti, ippopotami, bufali, giraffe, zebre, gnu, leoni, leopardi, coccodrilli, iene, facoceri e porcospini, senza contare una varietà di antilopi, tigri dai denti a sciabola e primati quali babbuini e colobi.

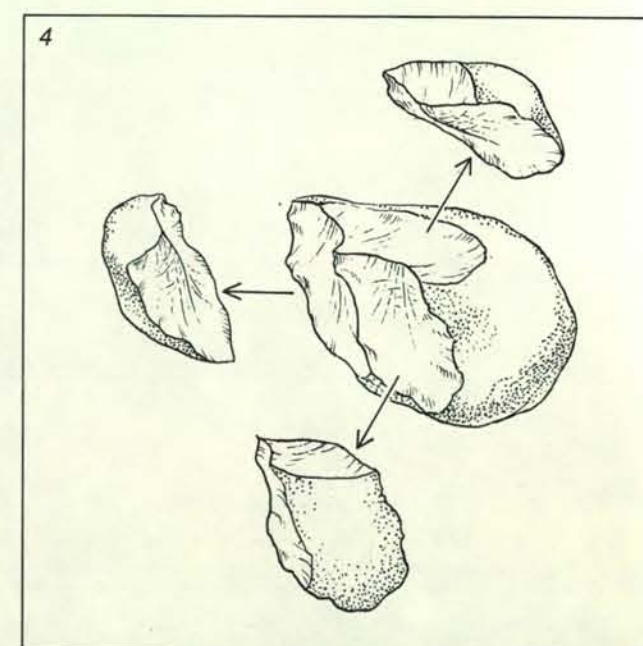
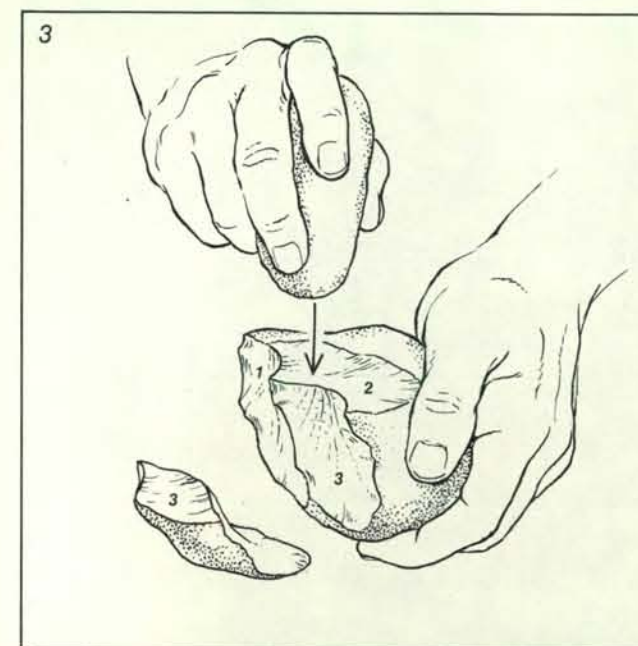
Australopiteci e primi uomini

I ritrovamenti fossili indicano che in questa zona il medesimo paesaggio era abitato da due diverse forme di ominidi. Entrambe le specie costituivano esperimenti di locomozione bipede nell'ambito delle grandi scimmie: camminavano in piedi sugli arti posteriori. La prima specie, nota come *Australopithecus boisei*, era dotata di muscolatura robusta e di cervello piccolo. Gli australopiteci recenti, il gruppo che comprende la forma *boisei* e ominidi affini con cervello ridotto, erano probabilmente specializzati per una alimentazione basata su vegetali piuttosto coriacei. Sembrano indicarlo i loro denti molari di grandi dimensioni, i muscoli masticatori e le forme di usura dentaria.

L'altra scimmia bipede, cui è stato dato il nome di *Homo habilis*, era con ogni probabilità l'antenato diretto degli esseri umani. In confronto ai cugini australopiteci, *Homo habilis* appariva più slanciato (il termine tecnico è «gracile») e dotato di cervello più grande. Intorno a un milione e 600 mila anni fa questa specie sembra avere dato origine a un discendente chiamato *Homo erectus*, l'antenato immediato della nostra specie o *Homo sapiens*. Le caratteristiche anatomiche suggeriscono che entrambe le forme più arcaiche del genere *Homo* fosse-

ro non specializzate dal punto di vista alimentare: mangiavano probabilmente più tipi di cibo che non gli australopiteci, inclusa una notevole dose di cibi ipernutrienti come le sostanze animali.

Per quasi un milione di anni, fin verso un milione e 200 000 anni fa, gli australopiteci vissero nello stesso paesaggio africano occupato dalla linea *Homo*, tant'è vero che i loro resti scheletrici si trovano talvolta nelle medesime località. Dal punto di vista della tecnologia e degli adattamenti evolutivi il periodo della loro coesistenza è di fondamentale interesse. I più antichi manufatti datati in modo attendibile sono stati segnalati nella valle del fiume Omo in Etiopia e risalgono a circa 2 milioni e 400 000 anni fa. Tuttavia, se nei siti dell'Omo gli archeologi hanno trovato frammenti di ciottoli quarzosi e di altre pietre intenzionalmente spaccate, purtroppo non si è conservato con le pietre alcun osso animale e quindi non si può stabilire a quale uso i manufatti fossero destinati.

La lacuna è colmata in luoghi come Koobi Fora, appunto, o la gola di Olduvai in Tanzania. Queste località hanno fornito non soltanto manufatti litici elementari, risalenti a 1,8 o 1,9 milioni di anni fa, ma anche ossa fossili ben conservate. La tecnologia litica di questo momento è generalmente chiamata oldowana, con allusione ai siti di Olduvai, dove tali manufatti sono stati inizialmente descritti da Louis e Mary Leakey. La maggior parte degli arnesi oldowani è composta di ciottoli di roccia effusiva, o di blocchi di quarzo e quarzite, i quali,

La scheggiatura bifacciale era uno dei metodi principali usati dai primi fabbricanti di utensili. Di solito essa era eseguita con la tecnica della percussione con martello duro. Una pietra, fungente da martello, era tenuta in una mano, mentre la pietra da scheggiare (o «nucleo») era tenuta nell'altra (*1*). Asportata una scheggia, il nucleo era voltato in modo che la cicatrice della scheggia servisse da piano di percussione per ricavare la scheggia successiva (*2*). Per una scheggiatura efficiente, l'angolo tra il piano di percussione e la superficie esterna del nucleo non deve essere superiore a 90 gradi (*particolare*). Se l'asportazione di schegge continua (*3*, *4*), si perviene a un nucleo con margine bifacciale.

Arnesi di stile oldowano sono stati fabbricati dall'autore per replicare l'intera serie delle forme di nucleo e l'assortimento di schegge. Secondo la terminologia tradizionale le forme di nucleo della fila in alto sono (*da sinistra a destra*) percussore o martello, *chopper* unifacciale, *chopper* bifacciale, poliedro, raschiatoio su nucleo e discoide. Nella fila inferiore si possono notare un raschiatoio su scheggia (*a sinistra*) e sei schegge non ritoccate. L'autore ha concluso che nella più remota tecnologia litica le schegge devono essere state importanti quanto gli arnesi su nucleo.

Gli arnesi di stile acheuliano realizzati dall'autore, efficienti nella macellazione di grandi animali, comprendono un'ascia da pugno di forma ovata, un'ascia appuntita, una mannaia o *cleaver* e un picco (*fila in alto*), uno sferoide, un raschiatoio su scheggia e tre schegge (*in basso*).

percossi con un ciottolo come martello, hanno dato origine a un «nucleo» e a una serie di schegge o frammenti. Talvolta, le schegge venivano ritoccate o perfino trasformate a loro volta in nuclei.

A partire da circa un milione e 500 000 anni fa una nuova tecnologia si aggiunse al repertorio oldowano: essa va sotto il nome di tecnica acheuliana. Il suo marchio distintivo consiste nel fatto che la materia prima per la fabbricazione di uno strumento non era tanto il ciottolo di fiume, quanto lo scheggione ricavato da un blocco. Scheggioni di questo genere, lunghi non di rado più di 15 centimetri, offrivano un eccellente punto di partenza per fabbricare i grandi arnesi bifacciali che gli studiosi di preistoria chiamano picchi, asce da pugno o mannaie (*cleavers*).

Poiché le forme *Homo* e gli australopiteci risultano essere coesistiti per la maggior parte di questa trasformazione tecnologica, non è per ora possibile indicare con precisione chi sia stato il fabbricante della maggior parte dei più antichi manufatti. Indizi indiretti, tuttavia, suggeriscono che il principale fabbricante fosse *Homo*. Arnesi di pietra non sono mai associati direttamente con gli australopiteci che vissero prima della comparsa del genere *Homo*. La fabbricazione di arnesi di pietra continuò del resto a fiorire dopo l'estinzione della linea australopitecina, un milione circa di anni fa. Altra considerazione: quando lo stesso livello geologico dà manufatti litici insieme a fossili di ominidi, almeno alcuni degli ominidi appartengono a *Homo*. Infine, i rappresentanti del genere *Homo* erano dotati di cervello relativamente grande, una caratteristica che ne fa - presumibilmente - gli abituali fabbricanti di utensili.

La più antica scheggiatura della pietra

Nell'avviare la mia ricerca, mi sono proposto di comprendere la natura delle più antiche tecnologie della pietra mediante una impostazione sperimentale di tipo olistico, ossia globale. La tecnologia litica andava studiata come un sistema, riconoscendovi più fasi. La prima fase corrisponde all'acquisizione della materia prima; poi vengono la fabbricazione, l'utilizzazione e l'abbandono, e per ultimo il seppellimento geologico dei manufatti. Un'altra fase, il trasporto della pietra, poteva avere luogo in qualsiasi momento tra l'acquisizione della materia prima e il successivo abbandono della pietra scheggiata. Il mio intendimento era quello di esaminare l'una dopo l'altra queste fasi del sistema. Poiché la maggior parte dei siti di Koobi Fora ha fornito manufatti di stile oldowano, ho concentrato la mia ricerca su questo stile, senza ignorare peraltro la tecnologia acheuliana.

Per quanto concerne l'ottenimento della pietra per la confezione di utensili, un fatto importante da considerare è quanto distante gli ominidi trasportassero la materia prima, avanti la fabbricazione degli arnesi. Nei siti di Koobi Fora, che si trovavano lungo i corsi d'acqua al momento della frequentazione, la fonte primaria di materiale doveva essere rappresentata dai ghiaieti vicini. In questi casi la distanza di trasporto poteva essere di pochi metri. Ma in altri siti, soprattutto quelli sui delta fluviali o presso le sponde dell'antico lago Turkana, le ricerche geologiche hanno mostrato che le più vicine fonti di roccia idonea alla scheggiatura dovevano trovarsi ad almeno alcuni chilometri di distanza. Molti siti contemporanei nella gola di Olduvai forniscono prova di distanze di trasporto comparabili.

A scopo di confronto ho esaminato sia le ghiaie dei greti attuali sia quelle esposte in superficie al tempo delle antiche frequentazioni ominidi. Intendevo individuare quali fossero i tipi di roccia di cui disponevano i fabbricanti primitivi. Tanto nei campioni attuali quanto in quelli preistorici il litotipo prevalente è una roccia lavica scura. Di solito, più del 90 per cento dei ciottoli di un greto fluviale è formato da questa roccia, mentre il resto è ignimbrite, selce e quarzo. Sebbene le rocce più rare tendano a essere più facilmente scheggiabili e a dare margini più taglienti della roccia lavica, i più antichi ominidi di Koobi Fora non sembrano avere fatto alcuno sforzo per sceglierle. I tipi di roccia adottati per fabbricare arnesi sono direttamente proporzionali alla loro abbondanza nelle ghiaie preistoriche. A Olduvai, per inciso, vi sono indizi di una selettività leggermente maggiore.

Eppure gli artigiani protoumani di Koobi Fora erano perfettamente consapevoli di altre caratteristiche del loro materiale. Avevano un occhio attento quando si trattava di evitare i ciottoli lavici difettosi, meno idonei alla scheggiatura. La parte interna di un ciottolo fortemente alterato può essere ossidata a tal punto che la pietra, quando la si lavora, si scheggia in modo imprevedibile. Il segnale dell'ossidazione eccessiva è una fratturina superficiale dello spessore di un capello, la «fessura di alterazione». Le pietre con fessure di alterazione sono comuni fra i ciottolami torrentizi di Koobi Fora, ma i caratteristici frammenti che esse producono sono rari fra i manufatti qui scoperti. Risulta chiaro che i primi ominidi avevano già imparato a rifiutare il materiale di qualità inferiore.

La seconda fase del sistema è la fabbricazione dell'utensile. Nel corso di migliaia di esperimenti ho riprovato i medesimi materiali e procedimenti che debbono essere stati usati dagli antichi ominidi. Ho sperimentato tutta una serie di metodi di scheggiatura per vedere quali fossero quelli usati a Koobi Fora. Cito la percussione con martello duro (percuotere un nucleo con un colpo brusco e secco di percussore di pietra), la tecnica dell'incudine (percuotere il nucleo contro una pietra-incudine fissa), e la tecnica bipolare (percuotere il nucleo con il martello di pietra tenendolo posato sull'incudine). Le forme di frattura su nuclei, schegge e frammenti che risultano dall'impiego di ciascuna tecnica sono state analizzate unitamente all'efficienza complessiva della tecnica. L'esame comparativo ha consentito di concludere che la tecnica fondamentalmente usata a Koobi Fora era la percussione con martello duro, che per i ciottoli di roccia lavica rappresenta la più efficiente delle tre ricordate.

I nuclei che derivano dalla percussione con martello duro sono classificati convenzionalmente in un certo numero di tipi morfologici, ai quali gli archeologi hanno dato nomi come *chopper* (fenditoi), discoidi, poliedri e raschiatoi su nucleo. Nella maggior parte degli studi tradizionali si è accettata la tacita premessa che per gli artefici primitivi fossero questi nuclei gli arnesi fondamentali, laddove le schegge e i frammenti avrebbero rappresentato, in sostanza, gli scarti di lavorazione. Non solo, si è anche pensato che le forme di questi primitivi arnesi su nucleo corrispondessero ai «modelli mentali» degli antichissimi fabbricanti: forme geometriche coerenti in base alle quali essi effettuavano la confezione dell'utensile.

Mano destra o sinistra?

I risultati dei miei esperimenti indicano che si è dato troppo risalto ai nuclei a scapito delle schegge; è possibile che il rapporto tradizionale vada invertito: le schegge possono essere state gli arnesi primari e i nuclei - almeno in un gran numero di casi - soltanto il sottoprodotto della lavorazione. Gli esperimenti mostrano che molte delle primitive forme di nucleo possono nascere spontaneamente quando si staccano schegge o frammenti da un ciottolo, senza alcuna intenzione consapevole. Se è così, la forma di molti nuclei ricalca incidentalmente il processo di lavorazione senza fornire indicazioni circa l'intenzione dell'artefice né la funzione del manufatto. Le schegge con il margine affilato, al contrario, di solito risultano essere arnesi taglienti di estrema efficienza, come apparirà dalla descrizione degli esperimenti di utilizzazione da me compiuti.

L'idea che molte delle forme di nucleo oldowane non derivassero da modelli mentali è stata confermata mettendo al lavoro scheggiatori inesperti, «tagliapietra» privi di addestramento. Per capire meglio il processo secondo il quale un individuo impara a scheggiare la pietra, ho infatti integrato i miei esperimenti personali con alcune indagini di scheggiatura su campioni di persone senza alcuna preparazione relativa ai manufatti di pietra. Dapprima, questi tagliatori producevano per lo più nuclei grossolanamente scheggiati o logorati, difficilmente confrontabili con i manufatti

dell'Età della pietra antica. Tuttavia, è emerso che nel giro di poche ore la maggior parte dei novizi arrivava a padroneggiare la meccanica fondamentale della scheggiatura. Da questo punto in poi essi erano capaci di produrre l'intera serie delle forme di nucleo oldowane. Ancora più significativo è il fatto che molte di tali forme emergono spontaneamente nel corso dell'apprendimento della scheggiatura.

I più antichi manufatti litici di Koobi Fora non posseggono certo molta finezza tecnica, ma a me sembra chiaro che anche i più antichi ominidi erano arrivati a questo punto, il controllo delle conoscenze fondamentali. Per scheggiare la pietra con il martello duro in maniera controllata, occorre soddisfare contemporaneamente tre condizioni. Per prima cosa, il nucleo deve avere uno spigolo acuto (cioè uno spigolo con angolo inferiore a 90 gradi) presso il quale il martello possa battere. In secondo luogo, il nucleo deve ricevere un colpo brusco in un punto generalmente situato a circa un centimetro dallo spigolo acuto. Infine, il colpo deve essere indirizzato attraverso una zona relativamente massiccia, come una cresta o una gibbosità (*si veda l'illustrazione a pagina 105*). Lo studio dei manufatti di Koobi Fora rivela che i loro fabbricanti avevano acquistato l'abilità cognitiva e motoria indispensabile per fare fronte a tutte e tre le condizioni, producendo da un nucleo una determinata sequenza di schegge.

I principi che ho descritto possono essere applicati in almeno due tipi differenti di scheggiatura: quella bifacciale alterna e quella unifacciale. Nella scheggiatura bifacciale alterna, le schegge sono staccate alternativamente da facce opposte del nucleo, finendo per produrre due facce scheggiate. Nella scheggiatura unifacciale una serie di schegge è asportata da una singola faccia del nucleo. Le schegge prodotte con la tecnica bifacciale alterna non chiariscono se l'artefice lavorasse con la mano destra oppure con la sinistra. Ma nella scheggiatura unifacciale l'aspetto delle schegge dipende dalla tendenza a usare una mano più dell'altra: schegge «destre» indicanti un tagliatore destrimane, possono essere distinte da schegge «sinistre» in base alla forma del còrtice (*si veda l'illustrazione a pagina 110*). Il còrtice è lo strato di alterazione presente sulla superficie della pietra.

Questa possibilità di distinzione ha permesso alcune osservazioni affascinanti sui più remoti sviluppi del destrismo negli esseri umani. Ho analizzato il rapporto tra schegge destre e schegge sinistre nei miei esperimenti di scheggiatura, trovando il valore di 56:44. La cosa non mi ha stupito, in quanto io sono destrimane e quindi è normale che le schegge destre prevalgano. (La proporzione non è più alta perché gran parte della scheggiatura è bifacciale e quindi produce quantità uguali di schegge destre e sinistre.) Quando ho misurato il rapporto tra le schegge nelle collezioni di Koobi Fora, sono arrivato quasi esattamente allo stesso valore, 57:43. Anche altre collezioni di manufatti della più antica Età della pietra presentano questo rapporto. Questa scoperta stupefacente suggerisce che la preferenza per l'uso della mano destra si era già sviluppata fra gli esseri umani di 1,9-1,4 milioni di anni fa.

Mano e cervello

In un'ottica evolutiva la diffusa dominanza di una mano sull'altra sembra essere una caratteristica esclusiva degli esseri umani. Gli altri animali mostrano sì delle preferenze individuali per una mano o una zampa, ma in sostanza esse si bilanciano in modo tale che il rapporto tra uso dell'arto destro e sinistro tende ad aggirarsi intorno a 50:50 in qualsiasi popolazione. Fra gli esseri umani, invece, i destrimani prevalgono decisamente sui mancini in un rapporto di circa 90:10 e tale preferenza sembra dipendere da un controllo genetico. Non solo, ma questa «lateralità» manuale si associa a una marcata lateralizzazione del cervello umano. Nella maggior parte degli esseri umani odierni l'emisfero cerebrale sinistro (il controllore della dominanza della mano destra) costituisce il centro delle facoltà di concatenazione cronologica, fra le quali ricade anche la capacità di parlare. L'emisfero destro è più specializzato nell'organizzazione spaziale. La comparsa della preferenza per il destrismo negli ominidi di Koobi Fora può benissimo indicare lo sviluppo della lateralizzazione.

A sua volta la lateralizzazione cerebrale può avere portato con sé una accresciuta capacità di concettualizzare il futuro, una conclusione confermata dalle mie ricerche sul trasporto dei manufatti. Il trasporto era un argomento di mio particolare interesse. Lo avevo abbordato studiando le distanze di trasporto dalle fonti di materia prima, come si è visto, ma desideravo conoscere di più. Nell'affrontare il problema si è nuovamente dimostrato di grande aiuto il còrtice dei ciottoli lavici. Quando una pietra viene scheggiata, naturalmente i pezzi con il còrtice saltano via per primi. Ho esaminato la distribuzione di schegge corticali e non corticali prodotte dai miei esercizi di scheggiatura su diversi tipi di nuclei. Se gli ominidi avevano svolto l'intera successione degli atti di scheggiatura nello stesso sito, bisognava attendersi che vi fossero rappresentate determinate distribuzioni di schegge corticali e non corticali.

Se, invece, i nuclei erano stati portati da un luogo all'altro e lavorati un po' qui, un po' là, la distribuzione avrebbe dovuto risultare distorta: in alcuni siti dovrebbe esserci una quantità minore delle schegge corticali che si formano nei primi momenti della scheggiatura. È appunto ciò che ho trovato. Confrontati con i miei modelli della «popolazione» globale di schegge derivanti dalla fabbricazione di utensili, modelli elaborati con il calcolatore, i siti di Koobi Fora tendono ad avere percentuali eccessivamente elevate di schegge non corticali (quelle prodotte nelle ultime fasi della lavorazione). Si direbbe che gli ominidi primitivi portassero in giro con sé dei nuclei

Gli esperimenti di trattamento della carne hanno rivelato che arnesi litici diversi sono idonei a compiti diversi. Una scheggia grezza (*a sinistra*) si è rivelata funzionale nella spellatura di uno gnu, mentre il taglio di una zampa di una gazzella richiede piuttosto un'ascia da pugno bifacciale (*pagina a fronte a destra*). Per estrarre il midollo rompendo ossa lunghe di bovidi è particolarmente utile un ciottolo naturale oppure - come qui - un nucleo pesante (*a sinistra*); un nucleo simile permette di aprire un cranio di bovide per estrarne il cervello (*a destra*).

parzialmente scheggiati, forse tenendoli in contenitori rudimentali. Nei punti di sosta i nuclei potevano essere lavorati ancora un poco e in certe località di frequentazione abituale o prolungata, molto del materiale di scheggiatura restava a terra a formare quelle concentrazioni di manufatti che gli archeologi chiamano siti.

Queste distribuzioni nello spazio e nel tempo suggeriscono che gli ominidi della zona di Koobi Fora avessero capacità di organizzazione e di previsione notevolmente superiori a quelle degli attuali scimpanzé. Se alcuni scimpanzé sanno prepararsi degli arnesi per uso futuro, tuttavia il tempo che intercorre tra la confezione e l'utilizzazione supera di rado i pochi minuti. Inoltre gli scimpanzé, in circostanze normali, non trasportano oggetti per più di cento metri. Sembra invece che fra gli antichi ominidi i materiali grezzi potessero essere trasportati più o meno regolarmente in previsione di potenziali usi futuri. Lo sviluppo di simili facoltà mentali può essere stato fra i fattori fondamentali alla base dell'uso di arnesi. A sua volta, il crescente uso di arnesi può avere contribuito all'azione di nuove forze selettive che favorirono l'ulteriore evoluzione del cervello, come è stato suggerito da Sherwood L. Washburn dell'Università della California a Berkeley.

È molto probabile che un aspetto di questo ciclo di *feedback* - un ciclo di eccezionale importanza adattativa - sia stata la capacità di adoperare gli arnesi per una crescente varietà di scopi. In base alle attuali conoscenze a proposito dei cacciatori-raccoglitori contemporanei e dei primati non umani, ho compilato un elenco delle funzioni che i manufatti avrebbero potuto svolgere fra gli ominidi africani più arcaici. Ho quindi svolto esperimenti per verificare quali tipi di arnesi fossero i migliori per l'esecuzione di ciascuna funzione. L'elenco comprendeva operazioni come la macellazione di animali, la lavorazione del legno, la lavorazione delle pelli, la rottura di noci a guscio duro e lo scavo. Dalle risultanze di centinaia di esperimenti di fattibilità, sono giunto alle seguenti generalizzazioni.

La scheggiatura unifacciale produce schegge capaci di indicare con quale mano l'antico artefice normalmente lavorasse. Un individuo destrimane tiene il percussore nella mano destra, dominante, e il nucleo nella sinistra: durante l'asportazione di una serie di schegge da una faccia, il nucleo tende a essere ruotato in senso orario. Gli individui mancini invertono le funzioni e di solito ruotano il nucleo in senso antiorario. Il modo in cui il còrtice (o strato di alterazione) si distribuisce sulle schegge può rivelare quale mano tenesse il percussore. Le schegge con i numeri *2* e *3* in basso hanno orientazione «destra» e dovrebbero indicare un fabbricante destrimane. Sulle schegge prodotte da esecutori mancini, la lunetta di còrtice ha orientazione opposta. L'analisi di vaste collezioni di schegge preistoriche ha permesso all'autore di stabilire che il destrismo sembra essersi affermato fra gli ominidi primitivi tra 1,9 e 1,4 milioni di anni fa.

Le funzioni dei manufatti di pietra

Tutti i miei esperimenti di macellazione sono stati condotti su animali che erano morti per cause naturali o che erano stati uccisi a scopo alimentare con metodi tradizionali. I soggetti comprendevano specie domestiche, come capre, pecore, maiali, vacche e cavalli, e specie selvatiche come cervi, zebre, òrici, gnu e addirittura elefanti. Individuare l'arnese più efficiente dipende dal compito specifico a esso richiesto entro l'operazione complessiva della macellazione. Una scheggia grezza e affilata è estremamente efficiente nell'aprire la carcassa. È possibile smembrare e rimuovere facilmente la carne con una scheggia grezza o con una scheggia ritoccata provvista di un margine affilato. Per tagliare carne rinsecchita rimasta su una carcassa già depredata (un comportamento ominide che Lewis Binford dell'Università del New Mexico ritiene importante) l'ideale è uno scheggione o un nucleo con un margine tagliente. Per estrarre il midollo spaccando le ossa, o raggiungere il cervello aprendo il cranio, un ciottolo naturale o un nucleo pesante sono gli arnesi più idonei.

Anche il fabbricare bastoni da scavo o aste da caccia comporta l'uso di arnesi diversi: nuclei pesanti e affilati per staccare un ramo diritto da un albero, schegge e raschiatoi su scheggia per la rifinitura della punta, e una superficie litoide scabra per l'abrasione finale. Una serie di schegge semplici e ritoccate è lo strumentario più efficiente per raschiare la carne e il grasso da una pelle, o eliminare la pelliccia dalla superficie esterna. Martelli e incudini costituiti di ciottoli non modificati funzionano perfettamente nel rompere noci dure. Per scavare alla ricerca di acqua profonda o di vegetali sot-

Uno dei caratteri distintivi della tecnologia acheuliana è uno scheggione staccato da un blocco, che, intorno a un milione e mezzo di anni fa, gli ominidi cominciarono a trasformare in asce da pugno, picchi e mannaie, dei tipi mostrati nell'illustrazione in basso a pagina 106.

terranei come radici e tuberi, gli arnesi di pietra non sembrano essere così efficienti come quelli di altri materiali, come per esempio, nei miei esperimenti, corna intatte di antilope, ossa lunghe spaccate di grandi animali oppure bastoni da scavo appuntiti.

Il repertorio tecnologico dei più antichi gruppi di ominidi potrebbe avere incluso armi, quali aste di legno, clave, bastoni da lancio o ciottoli usati come proiettili. Nel mio studio non ho ucciso animali con armi primitive, ma ho preferito svolgere alcuni esperimenti destinati a chiarire il problema generale dell'uccisione. Aste di legno con punte di corno, spinte dentro la cavità toracica di capre e vacche morte, riescono a penetrare per non meno di 30 centimetri, tranne quando urtino contro una costola. Le stesse aste, nel caso di un elefante africano adulto, hanno potuto penetrare nel torace per soli tre o quattro centimetri. Bisogna tuttavia por mente al fatto che gli ominidi primitivi erano molto più forti degli esseri umani attuali di statura e costituzione media, ragion per cui queste semplici armi, nelle loro mani, possono essere state notevolmente più efficaci.

Fra gli utensili più elementari con cui ho fatto esperimenti posso menzionare numerosi tipi di contenitori. Semplici mezzi per trasportare acqua, cibo, o nuclei parzialmente lavorati potrebbero essere stati fabbricati con gusci di uova di struzzo, carapaci di tartaruga, pezzi di corteccia o di pelle animale. Un ramo usato come uno spiedo e gettato sulla spalla potrebbe essere stato un ottimo arnese per il trasporto della carne. A dispetto del loro aspetto modesto, contenitori del genere possono avere avuto un ruolo non trascurabile nell'evoluzione umana. Anzi, i primi utensili in ordine di tempo possono essere stati degli arnesi da trasporto che permisero a un gruppo di ominidi di portarsi dietro il cibo, attenuando la loro dipendenza dall'ambiente immediato.

Se gli studi funzionali sono preziosi nell'aiutarci a comprendere come i manufatti primitivi potrebbero essere stati usati, essi però non dimostrano come l'impiego realmente avvenisse. Due recenti innovazioni cominciano a colmare questa lacuna. La prima è lo studio microscopico delle caratteristiche levigature che si formano sui margini degli strumenti di pietra durante la lavorazione dei diversi materiali. Lawrence H. Keeley dell'Università dell'Illinois a Chicago Circle è un pioniere in questo tipo di studio (si veda l'articolo *La funzione degli utensili di selce nel Paleolitico* di L. H. Keeley, in «Le Scienze quaderni» n. 30, giugno 1986), e in collaborazione con lui ho potuto esaminare le microlevigature da uso su un campione di strumenti di Koobi Fora.

Tecnologia litica e risorse animali

Purtroppo l'analisi microscopica delle usure non è applicabile all'esame delle rocce laviche di grana media con cui sono solitamente fatti gli arnesi di Koobi Fora. Si tratta di una roccia troppo eterogenea e grossolana perché dia informazioni valide al microscopio e la stessa alterazione chimica può modificare la superificie di un manufatto lavico nel corso del tempo. Ma le rocce microgranulari e in primo luogo la selce, nonostante siano in numero minore, forniscono dati interessanti. Su 59 manufatti con spigoli acuti e inalterati sottoposti a esame da Keeley, nove recano levigature da uso. Quattro di essi conservano tracce riferibili a macellazione di animali, tre a lavorazione del legno, e due al taglio di vegetali non coriacei. Anche se il campione è esiguo, suggerisce comunque che verso un milione e mezzo di anni fa il trattamento della carne poteva essere diventato un'importante attività tecnologica, accompagnata dalla confezione di semplici arnesi di legno e dal taglio di erbe. Queste ultime potevano servire per posarvi sopra la carne e proteggerla così dallo sporco, durante lo smembramento di una carcassa, oppure, secondo un suggerimento avanzato da Keeley, per farne materiale da giaciglio.

L'altro tipo di progresso metodologico si basa sulla modificazione complementare: quella dell'osso sotto l'azione dello strumento. Striature su ossa fossili da Koobi Fora e dalla gola di Olduvai sono state riconosciute come le tracce provocate da ominidi, mentre con arnesi litici spellavano, smembravano e scarnificavano la salma. Queste ricerche, condotte da Henry T. Bunn dell'Università del Wisconsin a Madison, Richard Potts della Smithsonian Institution, e Patricia Shipman della Johns Hopkins University, corroborano la scoperta che il trattamento della carne era in quell'epoca un'attività significativa. Inoltre, in siti di entrambe le località, si sono trovate ossa lunghe, ricche di midollo, con tipi di frattura verosimilmente provocati da martelli litici. Come queste risorse animali fossero ottenute è materia di discussione - si dibatte l'importanza relativa della caccia e della depredazione di carogne - ma un fatto sembra ormai chiaro: la tecnologia basata sulla scheggiatura della pietra era parte integrante dello sfruttamento delle carcasse di animali.

Queste informazioni, unite ai miei prolungati studi sperimentali, mi hanno permesso di cominciare a delineare uno scenario di come sia apparsa la tecnologia della pietra scheggiata. Io penso che l'impulso decisivo per lo sviluppo degli arnesi di pietra scheggiata sia stato il passaggio a uno sfruttamento delle risorse alimentari animali (mediante caccia su piccola scala o depredazione di carogne), nell'ambito di un modo di approvvigionamento opportunistico ancora dominato dai cibi vegetali. Questa transizione alimentare ebbe luogo probabilmente più di due milioni di anni fa e potrebbe essere posta in relazione con mutamenti climatici di vasta portata. Intorno a due milioni e mezzo di anni fa il clima africano sembra essere diventato notevolmente più arido. Elisabeth S. Vrba della Yale University ha suggerito che il mutamento climatico possa avere causato estinzioni su larga scala di specie animali, ma anche un episodio di speciazione segnato dalla comparsa di nuove forme.

Una tecnologia basata sulla scheggiatura della pietra può avere offerto agli ominidi arcaici un mezzo di adattamento alle nuove condizioni. Con essa gli ominidi si inserirono opportunisticamente in nicchie ecologiche che erano state dominio di animali predatori o che si nutrivano di carogne. Gli ominidi, sul piano biologico, non sono molto equipaggiati per tali nicchie: canini piccoli, molari piatti e assenza di artigli fanno di noi dei carnivori inefficienti. Inoltre - osserva Keeley - pietre o altri materiali dotati di margini taglienti non si trovano facilmente in natura. Come gli scimpanzé attuali, gli antichi ominidi dovevano già possedere una tecnologia per la manipolazione di vegetali, basata su percussori litici non modificati. La rottura accidentale dei «martelli» o delle «incudini» di pietra può avere dato origine a quei frammenti affilati il cui potenziale per il taglio fu poi riconosciuto con prove ed errori.

In tal modo i nostri antenati riuscirono a integrare il loro equipaggiamento biologico. Le pietre taglienti divennero per essi l'equivalente dei canini e dei denti ferini, i denti atti a lacerare la carne propri dei carnivori, mentre le pietre più pesanti sostituirono le mascelle e mandibole capaci di triturare l'osso. Che le schegge siano state una parte fondamentale di questo processo è indicato da una scoperta fatta nel corso della collaborazione con Keeley: tutti e nove i manufatti di Koobi Fora che portano tracce di uso sono schegge semplici. Nessuno di essi sarebbe stato considerato uno «strumento» nella maggior parte delle classificazioni tradizionali dei tipi oldowani. Allo stesso modo, molte delle fini striature parallele scoperte sulle ossa di siti oldowani sembrano dover essere intese come segni lasciati da schegge grezze. E una volta assimilata questa nozione dello scheggiare la pietra, dovette probabilmente mettersi in moto un *feedback* decisivo tra l'uso di utensili e l'espansione del cervello.

Un ulteriore salto

Poi, verso un milione e mezzo di anni fa, venne un secondo salto tecnologico, coincidente con la comparsa dei manufatti di stile acheuliano. Le mie ricerche portano a concludere che molte delle «asce da pugno», «mannaie» o «picchi» della tecnologia acheuliana, sono in realtà arnesi eccellenti per la macellazione «pesante» dei mammiferi più grossi. Sembra dunque probabile che questo ulteriore progresso sia da associare alla macellazione sempre più sistematica di carcasse più grandi, ottenute con lo sfruttamento di vittime altrui o con la caccia, come suggerito da Clark e da altri in anni passati.

Il lavoro mio e di altri studiosi mostra così che le trasformazioni nella dieta e nella biologia e il più antico sviluppo tecnologico sono strettamente associati. Ricerche future aumenteranno senza dubbio le conoscenze e potranno modificare le nostre interpretazioni. Ciò che non muterà è l'intrinseco fascino degli elementari manufatti di pietra scoperti nei più antichi siti africani. Essi rappresentano la prima testimonianza di cultura (ossia della trasmissione di comportamenti appresi) nella documentazione preistorica. Cultura e tecnologia, nelle loro complesse manifestazioni a noi contemporanee, sono spesso date per scontate, fino a perdere la nozione delle loro origini. Eppure ogni tecnica moderna, incluse quelle che ci permettono di esplorare altre parti del sistema solare, di scindere l'atomo, di sviluppare un'ingegneria genetica o di eseguire trapianti di organi, è collegata da un filo ininterrotto di cultura ai primordi della fabbricazione di utensili di pietra scheggiata, tra due e tre milioni di anni fa.

(RI)CREAZIONI AL CALCOLATORE

di A. K. Dewdney

Anche il suono del calcolo può essere musica

Era il 1965 quando sentii per la prima volta il suono del calcolo. Al centro di calcolo dell'Università del Michigan, un IBM 7090 era stato dotato di un semplice trasduttore elettromagnetico collegato a un altoparlante. Mentre un programma girava, un particolare registro della macchina cambiava il suo contenuto molte migliaia di volte al secondo. L'incredibile accozzaglia di suoni alieni che risultava dalla serie di minuscoli scatti era tutto un alternarsi di ronzii, strilli, brontolii e lamenti. Qualche volta, un suono stridente passava dal grave all'acuto; si trattava forse del suono prodotto da un doppio ciclo del programma; il ciclo interno, eseguito sempre più velocemente, creava un suono di altezza crescente.

L'esperienza ha suggerito una grande varietà di programmi che esplorano diversi aspetti della melodia, dell'armonia e del ritmo. Anche se i calcolatori domestici non dispongono normalmente di trasduttori elettromagnetici comandati da registri accumulatori, la maggior parte di essi ha piccoli altoparlanti collegati a primitivi generatori di suoni. In molti fra i più comuni linguaggi di programmazione vi sono istruzioni che, quando vengono eseguite, generano una grande varietà di suoni. Questo semplice strumento può essere sfruttato sia per rendere udibili dei programmi, sia per produrre programmi fatti apposta per l'ascolto. Nel primo caso, si aggiunge un'istruzione per la generazione del suono in un programma originariamente finalizzato al calcolo. Di questo argomento parlerò brevemente più avanti. Nel secondo caso, si costruisce deliberatamente un programma perché produca effetti melodici, armonici o ritmici. Due dei programmi descritti in questo articolo si ricollegano al primo tipo: alcune istruzioni per la generazione del suono sono inserite in un codice per altri versi normale.

È facile generare con un programma una melodia, intesa come semplice successione di note. In effetti, il programma può essere formato da un unico ciclo. Un processo decisionale inserito nel ciclo stabilisce volta per volta quale nota debba essere suonata e quale debba essere la sua durata. Il programma può essere poi complicato a piacere. A quanto mi risulta, esistono già programmi di questo tipo che generano melodie convincenti in un dato stile tradizionale. Personalmente sarei lieto di ricevere notizie da chi ha più esperienza. Nel frattempo, lascerò che a dar vita alla melodia siano i programmi già disponibili.

I più umili tra i candidati al ruolo di produttori di melodie sono semplici algoritmi aritmetici. È facile trasformare in note l'uscita di questi algoritmi attraverso un'enorme varietà di possibili codificazioni. La più semplice utilizza l'assegnamento lineare congruente, un procedimento più breve del suo stesso nome:

$$x \leftarrow (a \times x + b) \bmod m$$

Se si specificano in anticipo i parametri a, b e m, un valore iniziale della variabile x è trasformato in una successione di valori attraverso la continua iterazione dell'assegnamento. L'espressione «mod m» è un'abbreviazione di «modulo m»: significa che il numero calcolato tra parentesi deve essere trattato come le ore di un orologio con m ore. Per esempio, 10 modulo 8 è uguale a 2. Quindi se m è 8 e se a, b e il valore iniziale di x sono tutti interi, si ottiene una successione di numeri che variano da 0 a 7.

Mediante una semplice tabella è facile trasformare la successione di numeri in una successione di note:

0	1	2	3	4	5	6	7
do	*re*	*mi*	*fa*	*sol*	*la*	*si*	*do*

Nel programma che chiamo SOLFEGGIO, le grandezze simbolizzate dai nomi delle note sono sostituite dalle frequenze, espresse in cicli al secondo, della scala di do maggiore che inizia con il do centrale:

0	1	2	3	4	5	6	7
262	294	330	349	392	440	494	523

L'intero programma può essere riassunto nella consueta forma algoritmica:

```
input a, b, x
per i = 1 a 100
    x ← (a × x + b) mod 8
    nota ← note(x)
    suona nota
```

Nella mia versione di SOLFEGGIO, i numeri della tabella chiamata *note* sono utilizzati direttamente dall'istruzione che genera una nota dell'altezza appropriata. Molti linguaggi consentono anche il controllo della durata della nota, ma per ora sarà sufficiente fissare la durata a, diciamo, mezzo secondo.

A questo punto si apre tutto un universo melodico a chi abbia «orecchio» per l'aritmetica. A seconda dei valori scelti per i tre parametri, si possono sentire noiose monodie di staccati, bizzarre melodie ripetitive su poche note o una strana musica tempestosa piena di sbalzi e improvvise scale. A proposito di quest'ultima categoria, mi è capitato a volte di essere preso alla sprovvista da qualche delizioso piccolo pezzo musicale, perduto per sempre perché non avevo preso nota dei parametri che l'avevano generato. L'algoritmo che ho proposto produce melodie lunghe 100 note, lunghezza che può essere ridotta da chi voglia esplorare un numero maggiore delle possibilità offerte da questo nuovo terreno musicale.

SOLFEGGIO può e deve essere migliorato. Si può scegliere un valore più grande per il modulo m, grande abbastanza da abbracciare due o più ottave diatoniche. Non c'è poi alcuna restrizione modale; invece delle scale diatoniche si possono scegliere scale dodecafoniche comprendenti tutti i semitoni dell'ottava. Si può persino scegliere la forma ultramoderna di musica lineare congruente in cui i numeri x prodotti dall'algoritmo specificano frequenze in modo più diretto attraverso l'aggiunta di una costante, per esempio 100. Se a un certo punto x è 183, allora 283 è la frequenza della nota che il programma suonerà. Mi chiedo quali altre possibilità possa inventare il lettore. Nel frattempo ecco un'interessante domanda sulla monotonia. Per dati valori di a, b e m, quante note verranno suonate prima che la melodia inizi a ripetersi?

Nota	DO	DO#	RE	RE#	MI	FA	FA#	SOL	SOL#	LA	LA#	SI	DO
Frequenza	261,6	277,2	293,7	311,1	329,6	349,2	370,0	392,0	415,3	440,0	466,2	493,9	523,3

Le frequenze delle note sopra o sotto questa ottava si ottengono moltiplicando o dividendo per 1,05946 e arrotondando in modo adeguato. Il numero è $\sqrt[12]{2}$, la radice dodicesima di due.

Frequenze della scala di semitoni dal do centrale a un'ottava sopra il do centrale

L'armonia rientra attualmente nella competenza vocale di molti calcolatori domestici. Quelli dotati di due o più altoparlanti saranno in grado di suonare l'intero repertorio generato da un programma che chiamo CANONE. Ci saranno anzi degli appassionati che riusciranno senza dubbio a migliorare il progetto. Alcuni lettori avranno già acquistato una MIDI, acronimo di *Musical Instrument Digital Interface* (interfaccia digitale per strumenti musicali), una scatola nera elettronica che traduce segnali generati da un calcolatore in comandi per strumenti elettronici quali tastiere musicali e singoli canali di sintetizzatori. (I lettori che vogliano avere ulteriori notizie sulla MIDI possono scrivere alla International MIDI Association, 12439 Magnolia Boulevard, Suite 104, North Hollywood, California 91607.)

CANONE genera armonie a due voci con linee melodiche quasi identiche. Esempi di questo tipo, se non della specie, sono canoni quali «Row, Row, Row Your Boat» e «Fra' Martino». CANONE genera un canone noto nella tradizione accademica come imitazione di prima specie, che è il primo passo nello studio del contrappunto.

Un'armonia di questo genere ha due linee melodiche che soddisfano quattro criteri. Primo, tutte le note hanno la stessa durata. Secondo, una linea inizia dopo l'altra. Terzo, le due linee sono identiche tranne per il fatto che una è trasposta verso l'alto di un certo intervallo convenzionale (unisono, quarta perfetta, quinta perfetta oppure ottava). Quarto, le due linee insieme devono soddisfare, nota per nota, determinate regole dell'imitazione di prima specie. Queste regole, tutte reperibili in testi classici, comprendono gli intervalli armonici consentiti tra le note (*si veda l'illustrazione nella pagina successiva*) e stabiliscono la forma delle linee melodiche. Quest'ultimo aspetto è stato qui tralasciato per ragioni di semplicità.

Nella figura in alto di pagina 120 si può vedere un esempio di imitazione di prima specie. Questo esempio è stato scritto da un uomo, non da una macchina. Basta un'occhiata per rendersi conto che a un certo punto il compositore si libera dai vincoli imposti da tutti e quattro i criteri. La regola tre di solito è la prima a saltare; è sufficiente che la seconda linea imiti la prima nello spirito.

Un programma per calcolatore, invece, rimane imperterrito di fronte alle limitazioni. CANONE continuerà a lavorare instancabilmente finché non avrà generato un intero pezzo. Di solito, comunque, il risultato del programma è noioso. Per evitare questo problema ho generato un gran numero di brevi canoni e ho catalogato i più interessanti, alcuni dei quali poi possono essere accostati a formare pezzi più lunghi (*si veda l'illustrazione a pagina 120 in basso*). L'esperimento è stato condotto dieci anni fa con l'assistenza di Gregory Utas, uno dei miei studenti all'Università dell'Ontario occidentale.

ROCK:

1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	PIATTO
0	0	0	0	1	0	0	1	0	1	0	0	1	0	0	1	RULLANTE
1	0	1	1	0	0	0	0	1	0	1	1	0	1	0	0	CASSA

REGGAE:

1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	CHARLESTON
0	1	0	0	1	0	0	1	0	1	0	0	0	0	0	0	RULLANTE
0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	CASSA

JAZZ:

1	0	0	1	0	1	1	0	0	1	0	1	PIATTO
0	1	1	0	1	0	0	1	1	0	1	0	RULLANTE
1	0	0	0	0	1	1	0	0	0	0	1	CASSA

SAMBA:

1	1	0	1	1	0	1	1	0	1	1	0	1	1	0	1	RULLANTE
0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	CERCHIO DEL RULLANTE
1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	CASSA
0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	CHARLESTON

Quattro esempi di ritmi per BEAT

CANONE procede nella creazione di un brano con un metodo di enumerazione. Supponiamo, per esempio, che gli sia stato detto di generare un canone in miniatura lungo sei note. Per ragioni che risulteranno subito ovvie, l'estensione coperta da ciascuna linea melodica è piccola, diciamo di sei semitoni verso l'alto e sei semitoni verso il basso rispetto alla tonica, ossia alla nota d'inizio. Le 13 note possibili possono essere viste come cifre di un sistema in base 13. La cifra 0 rappresenta la nota sei semitoni sotto la tonica e la «cifra» 13 rappresenta la nota sei semitoni sopra. CANONE procede semplicemente enumerando tutti i numeri di sei cifre in base 13. Ciascuno di essi, dopo tutto, codifica una melodia di un certo tipo.

Ogni volta che CANONE genera una nuova linea melodica ne crea una copia, la trasporta verso l'alto di una quinta e la sposta in avanti di un dato numero di note, per esempio due. Confronta poi le note che l'operazione ha portato in coincidenza temporale. Se nessuna delle coppie risultanti viola le regole dell'armonia, la linea è accettata e viene stampata direttamente per il compositore in attesa oppure viene conservata in un archivio prima di essere stampata.

In particolare, CANONE richiede che il compositore specifichi tre parametri prima di far girare il programma: *int*, l'intervallo di imitazione, *rit*, il ritardo dell'entrata della seconda linea melodica, e *num*, il numero di note di ciascuna linea. Le note della linea in via di generazione sono conservate in una matrice chiamata melodia, *mel*. Ecco uno schema algoritmico di CANONE:

```
input int, rit, num
mel(1) ← 7
per i = 2 a num
    mel(i) ← 1
trovato ← falso
mentre trovato falso
    incrementa mel
    per j = 1 a num − rit
        confronta mel(j) e
        mel(j + rit) + int
    se consonante allora
    trovato ← vero
output mel
opzione per continuare
```

L'algoritmo inizia assegnando a *mel*(1), la prima nota del canone, il valore 7. Questa è la tonica e non cambierà. Gli altri elementi di *mel* iniziano tutti con il valore 0. Un ciclo «mentre» (*while*) controlla una variabile booleana, cioè logi-

a

NOTA SUPERIORE (colonne) / NOTA INFERIORE (righe)

	DO	DO#	RE	RE#	MI	FA	FA#	SOL	SOL#	LA	LA#	SI	DO
DO		X	X			X	X				X	X	
DO#			X	X			X	X				X	X
RE				X	X			X	X				X
RE#					X	X			X	X			
MI						X	X			X	X		
FA							X	X			X	X	
FA#								X	X			X	X
SOL									X	X			X
SOL#										X	X		
LA											X	X	
LA#												X	X
SI													X
DO													

b

NOME	NUMERO DI SEMITONI
1	0
1+/2°	1
2	2
3⁻	3
3	4
P4	5
4+/5°	6
P5	7
6⁻	8
6	9
7⁻	10
7	11
8	12

Le note suonate insieme sono consonanti, secondo la tabella *a*, se le combinazioni non cadono su una X. Nella tabella *b* si trovano i nomi e le dimensioni (i semitoni) dei risultanti intervalli armonici. Gli intervalli di unisono (1), terza minore e maggiore (3⁻ e 3), quarta perfetta (P4), quinta perfetta (P5) sesta minore e maggiore (6⁻ e 6) e ottava (8) sono considerati consonanti nell'imitazione di prima specie. I corrispondenti numeri di semitoni sono 0, 3, 4, 5, 7, 8, 9 e 12.

Tabella degli intervalli armonici consonanti

ca, chiamata *trovato* a cui inizialmente è assegnato valore falso. *Trovato* diventa vera se le istruzioni del corpo centrale del ciclo scoprono una valida linea melodica canonica. Per prima cosa, viene incrementata la matrice *mel*. La matrice viene esaminata da destra verso sinistra e la procedura di enumerazione cerca un elemento che sia inferiore a 13. Quando lo trova, gli aggiunge 1 e mette a 0 tutti gli elementi sulla destra (se ce ne sono). Questo è proprio quello che avviene quando si conta nel modo consueto, con il 9 che sostituisce il 13. Per esempio, 3572 + 1 = 3573, 3579 + 1 = 3580 e 3599 + 1 = 3600.

Il successivo compito del ciclo consiste nel confrontare ogni nota *mel(j)* con la nota *mel(j + rit) + int*. In altri termini, il programma somma *int* alla nota che si trova *rit* note dopo *mel(j)* e cerca la differenza tra i valori delle due note nella tabella di regole per l'imitazione di prima specie. Se la differenza è considerata consonante in tutti i casi da 1 a *num − rit*, la variabile booleana *trovato* è posta uguale a vero. (Le note che si trovano al di là di *num − rit* sono suonate da sole, senza contrappunto, quindi non ci sono differenze da cercare nella tabella.) Una volta trovata e stampata una simile linea melodica, il programma chiede al compositore «Vuoi continuare?». Se la risposta è sì, il programma torna all'istruzione *trovato* ← falso e riprende da dove si era interrotto.

Naturalmente, è possibile suonare le linee melodiche scoperte da CANONE con i piccoli altoparlanti del proprio calcolatore domestico. I lettori con inclinazioni musicali, invece, saranno in grado di canticchiare la linea melodica oppure di trascriverla su pentagramma. A quel punto, il canone può essere eseguito su un'altra tastiera in tutto il suo fasto armonico.

Il ritmo è un aspetto della musica più complesso di quello che probabilmente sono in molti a supporre. Tradizionalmente, la musica occidentale non è mai stata molto elaborata dal punto di vista ritmico, mentre la cultura musicale popolare ha toccato una straordinaria varietà di forme ritmiche. Molte di queste hanno avuto origine, direttamente o indirettamente, dalla musica tradizionale africana o asiatica, compresa la maggior parte della musica rock, il jazz, la musica caraibica e quella latino-americana. Gli occidentali sono anche sempre più consapevoli del complesso contribuito delle tabla (una coppia di tamburi che si suonano con le dita) alle forme musicali indiane come il raga.

Il programma che chiamo BEAT consente di tradurre semplici ritmi in successioni di 0 e 1, i quali a loro volta sono trasformati in suoni attraverso il semplice espediente di percorrere più volte ciclicamente la successione. A ogni passaggio, un 1 provoca la breve emissione di un suono, uno 0 niente.

In effetti, BEAT è sufficientemente semplice per poterlo descrivere senza ulteriori spiegazioni. Da un punto di vista strutturale, è abbastanza simile a SOLFEGGIO (il programma che suona musica lineare congruente). Una singola matrice *impulso* tiene il ritmo conformemente a quanto è stato specificato dal programmatore all'inizio dell'esecuzione:

```
input impulso, num, dur
per i = 1 a 25
  per j = 1 a num
  k ← 1
  mentre k ≤ dur
    k ← k + 1
  se impulso(j) = 1
    allora suona
```

Le variabili chiamate *num* e *dur* si riferiscono rispettivamente alla dimensione della matrice d'ingresso e alla durata delle pause tra i suoni. Il ciclo esterno specifica che l'intervallo ritmico di base determinato da *impulso* sarà suonato 25 volte. Questo numero può essere facilmente modificato dai lettori che volessero ascoltare per un periodo di tempo più lungo un certo ritmo generato dal programma. Il successivo ciclo interno controlla l'indice della matrice; l'algoritmo considererà di volta in volta ciascun elemento per decidere se produrre o meno un suono. Che intervallo temporale ci deve essere tra gli eventi sonori? Questo è stabilito da uno speciale ciclo d'attesa che si limita a contare fino alla durata specifica, *dur*. Poi, se *impulso(j)* è 1, il programma BEAT produrrà una nota, emetterà un ronzio o qualsiasi altra cosa disponibile. Se *dur* è piccolo, il ritmo sarà veloce; se *dur* è grande, il ritmo sarà lento.

Anche BEAT, come i programmi precedenti, può essere migliorato. Si può ottenere qualcosa di simile a un complesso di percussioni impiegando suoni diversi (preferibilmente quelli bassi che rappresentano vari tipi di tamburi, campane o piatti). Anche se i suoni saranno forse lungi dall'essere realistici, i ritmi saranno soddisfacenti. Si può complicare la matrice *impulso* usando interi quali 0, 1, 2 e 3 per rappresentare rispettivamente il silenzio, un tamburo alto che suoni da solo, un tamburo basso che suoni da solo e i due tamburi che suonino insieme. In ciascuno di questi casi, l'elemento di *impulso* in via di esame nel ciclo interno deve essere decodificato da un insieme di istruzioni condizionali che controllino l'esecuzione del silenzio, di un suono o di entrambi i suoni. In quest'ultimo caso, la simultaneità è resa con l'esecuzione consecutiva di due note. Se si trova che l'esecuzione di due suoni consecutivi rende impossibile la sincronizzazione, allora il ciclo di attesa deve essere spostato negli enunciati condizionali in modo che ci sia una copia per ciascuna delle quattro possibilità. Nel caso del silenzio e dei due suoni, il limite *dur* deve essere sostituito da nuovi limiti, uno più lungo e uno più corto di *dur*. In entrambi i casi, la differenza sarebbe il valore temporale assegnato a una singola nota.

Una versione più raffinata di BEAT sarà senza dubbio in grado di suonare alcuni degli interessanti ritmi illustrati nella pagina precedente. I diversi spartiti sono stati interpretati come successioni di eventi discreti collegati su una base temporale comune e facilmente tradotti nel contenuto di *impulso* attraverso codificazioni predisposte dall'aspirante compositore.

Al di là delle incursioni effettuate nel campo della melodia, dell'armonia e del ritmo, continuo a essere affascinato dall'idea di affidare due ruoli a ogni programma per calcolatore. Nel primo ruolo il programma calcola ciò che deve calcolare. Nel secondo ruolo, nel programma si associano una nota o due ai suoi cicli interni, ai cicli esterni e alle istruzioni condizionali. Il programma ha una canzone da cantare per ogni problema che gli venga sottoposto. Chi ascolti con regolarità il proprio programma preferito (sia esso ricreativo o commerciale), si farà un orecchio per la sua esecuzione. Non ho dubbi che certi errori in nuove versioni di un programma possano essere individuati a orecchio. I lettori che abbiano idee in merito all'argomento sono pregati di sottopormele.

Nell'articolo del mese scorso sui veicoli di Valentino Braitenberg e sui loro cervelli elettronici, ho proposto un rompicapo in cui compariva una sfera dorata lasciata cadere da un UFO; una volta attivata, la sfera rotola su un pavimento liscio emettendo un ronzio. Come fa a muoversi se non ha appendici esterne quali ruote o gambe? Inoltre, la sfera evita qualunque ostacolo incontri, rotolandoci intorno. In che modo può riuscirvi?

Molti anni fa, comparve per qualche mese nei negozi americani un giocattolo chiamato Moon Walker, ossia «veicolo lunare». Era formato da due emisferi di plastica cavi e trasparenti, che si univano tra loro racchiudendo un'automobilina a batteria. Per far funzionare il giocattolo, si accendeva il motore elettrico dell'automobilina prima di mettere il veicolo in uno degli emisferi. La sfera di 10 centimetri che ne risultava rotolava sul pavimento perché l'automobile all'interno cercava continuamente di scalare la parete interna della sfera nella direzione in cui il veicolo si trovava a essere puntato. La sfera rotolava nella direzione del conseguente squilibrio.

Quando la sfera incontrava un ostacolo come la gamba di una sedia, l'automobilina si arrampicava ancora più in alto nella sfera. A quel punto entrava in azione sulla ruota motrice destra un sistema incorporato di slittamento. L'automobile continuava a sbandare verso destra, muovendosi sulla superficie interna fino a puntare nella direzione in cui la sfera non incontrava resistenza. In questo modo la sfera poteva aggirare gli ostacoli. La perplessità degli osservatori era massima quando la sfera veniva dipinta in modo da nascondere l'automobilina. A quel punto non c'era limite a ciò che si poteva immaginare sul contenuto della misteriosa sfera venuta dallo spazio.

Da quando nell'articolo di marzo ho parlato del primo torneo internazionale di Guerra dei nuclei, il numero dei membri della International Core Wars Society è quadruplicato, a quanto dice il suo direttore Mark Clarkson. Chi fosse interessato può scrivere a Clarkson, 8619 Wassall Street, Wichita, Kansas 67210. Anche William R. Buckley, direttore di «The Core War Newsletter», ha reso noto che si è verificata una forte crescita del numero di abbonamenti. Buckley spiega che la pubblicazione servirà da principale organo di comunicazione per i membri della società che vogliano esporre miglioramenti nelle regole, nuove tecniche di battaglia e nuovi livelli di gioco. Chiunque desideri abbonarsi al bollettino scriva a Buckley, 5712 Kern Drive, Huntington Beach, California 92649.

Esempio, tratto da un manuale, di imitazione di prima specie a due parti

A

B

AB

Due canoni generati dal calcolatore vengono uniti per formarne uno più lungo